University of Sussex

Semi-Analytic Model of Galaxy Formation with Radiative Feedback During the Epoch of Reionisation

Chaichalit Srisawat

Submitted for the degree of Doctor of Philosophy

University of Sussex

January 2016

UNIVERSITY OF SUSSEX

Chaichalit Srisawat, Doctor of Philosophy

Semi-Analytic Model of Galaxy Formation with Radiative Feedback During the Epoch of Reionisation

Summary

Several hundred million years after the Big Bang, the Epoch of Reionisation(EoR) started as the photons from the first objects ionised neutral baryons in the Universe. The observations such as the Gunn-Peterson troughs in quasar absorption spectra and the linear polarisation of the cosmic microwave background (CMB) impose strong constraints on reionisation models of the EoR. Recent data provide the rest-frame ultraviolet luminosity of galaxies up to redshift 10. However, the observation of star formations in low mass galaxies is still not practicable. Their star formations are expected to be suppressed by the increase of ionised baryons and greatly affect the reionisation models.

We develop a flexible pipeline which utilises the Munich Semi-Analytic Model of galaxy formation, L-Galaxies, and a semi-numerical modelling of cosmic reionisation. This combination allows us to create a self-consistent reionisation simulation in computational models of galaxy formation. We use this pipeline on a high resolution cosmological N-body simulation to produce the redshift evolution of the star forming galaxies during the EoR. Comparisons of the properties of mock galaxies and the growth of ionised hydrogen bubbles suggest that the reionisation history heavily depends on the suppression models used in the modeling of dwarf galaxy formation.

During this research, some numerical flaws of merger tree generation algorithms were identified. We investigated the origins of these problems and present suggestions for solving them.

Acknowledgements

I would like to express my gratitude to my supervisors, Dr Ilian Iliev and Prof Peter Thomas, whose expertise, understanding, and saintly patience, added considerably to my PhD experience. I appreciate their extremely useful supervisions throughout the years, especially their tolerance to see the same plots nearly every week for almost two years until I noticed the mistakes. I also thank them for giving me opportunities to work with many people around the world. Without these works, this thesis would never be completed.

I would like to specially thank some of my collaborators; Dr Bruno Henriques for many useful helps with L-Galaxies; Dr Suman Majumdar for his semi-numerical code which is used in this research; Prof Paul Shapiro, for his TACC allocations which I used in this work, both with and without his knowing; Dr Frazer Pearce and Dr Alexander Knebe, if not for their slave driving techniques, many of my works could not have been done; Dr Jiaxin Han, without him, I would never be able to debug the codes in Sussing Merger Trees workshop.

Thanks also go out to my friends at Sussex; Dr William Watson for giving me the introduction about CubeP^{3}M; Scott Clay, for sharing HST observational data with me; David Sullivan for very helpful results from his simulations; Benoît Fournier, for his new infall code; Dr Keri Dixons and Hannah Ross for their helps with C^2-ray and the suppression models; Alexander Eggemeier, without his red pens, the proofreading would have been much more difficult. Above all other things, I need to thank them for playing with me in my very limited time at the department.

I need to give my thanks to P'Lek, Somsong Arunrat, who kept bringing me food when I had gout attacks and was unable to walk during this thesis was being written up. I am also thankful to P'Mai, Dr Jirapas Jongjitwimol, for his advices on Turnitin system and thesis bindings.

I would also like to thank my family for the support they provided me through my entire life and in particular, I must acknowledge my wife, Monticha, without whose love, encouragement and editing assistance, I would not have finished this thesis, for real.

In conclusion, I recognise that this research would not have been possible without the financial support of the Development and Promotion of Science and Technology Talents Project, and express my gratitude to the program.

Contents

List of Tables

List of Figures

Part I

Introductions

Chapter 1

Semi-Analytic Modelling of Galaxy Formation

1.1 ΛCDM Cosmology

The ΛCDM is a model where the dark matter is assumed to be non-relativistic at the epoch of matter-radiation equality. It is also assumed to be dissipationless and collisionless, which mean the dark matter cannot cool by radiative photons and can interact with other particles only through gravitational attraction respectively. In addition, the Universe is assumed to be flat, homogeneous and isotropic (in the large scales) according to the cosmological principle. The letter Λ represents the cosmological constant which is currently associated with the dark energy in empty space.

From the inflation model, the primordial perturbations have a nearly-scale-invariant power spectrum. This predicts that structures grow hierarchically. As gravity accumulates the dark matter in the over-density regions, the baryons fall into these potential wells after decoupling. As the result, the acoustic oscillations imprinted in the baryons are shown as a series of low amplitude peaks in the matter power spectrum.

The Big Bang model of cosmology became popular after the the discovery of the cosmic microwave background (CMB) in 1965. Cold Dark Matter models came under active consideration after COBE, discovered the CMB anisotropy in 1992 (Smoot et al., 1992). After Ia supernovae (SNs) reveal that the Universe is now undergoing an accelerated expansion (Perlmutter et al., 1999), the ΛCDM model became the leading model of the Universe. There are many other observations to support it afterwards from the start of the 21th century; the BOOMERanG microwave background experiment which measured the total density to be close to the critical density (Crill et al., 2003), 2dFGRS galaxy redshift

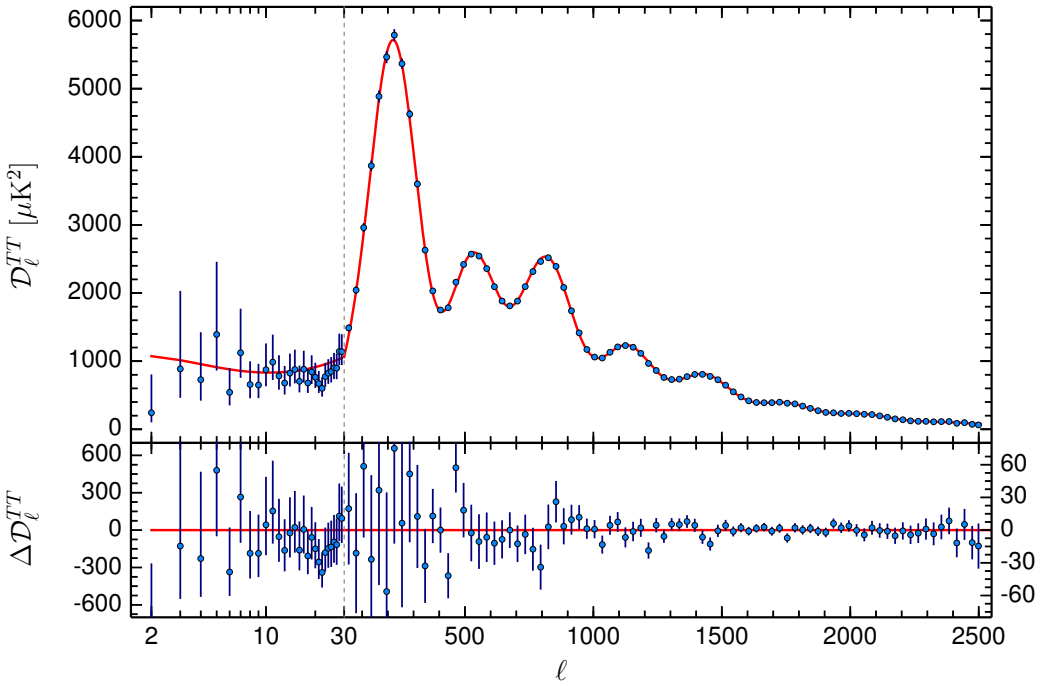

Figure 1.1: The *Planck* 2015 temperature power spectrum. The best-fit base CDM theoretical spectrum fitted to the *Planck* TT+lowP likelihood is plotted in the upper panel. Residuals with respect to this model are shown in the lower panel. The error bars show $\pm 1\sigma$ uncertainties (Planck Collaboration et al., 2015).

survey to measure the matter density close to 25% (Peacock et al., 2001). The precision cosmology observation of the microwave background from the successors of COBE, the Wilkinson Microwave Anisotropy Probe (WMAP), in 2003-2010 (Spergel et al., 2003, 2007; Komatsu et al., 2009, 2011a; Hinshaw et al., 2013) and *Planck* in 2013 - 2015 (Planck Collaboration et al., 2014, 2015) have continued to increase the precision of cosmological parameters to be under 1% uncertainty nowadays.

1.2 Cosmological simulation

For many decades, tremendous efforts have been put into studies to understand cosmic structures and galaxy formation. At the beginning, most of the understanding came from elegantly simple analytic arguments and insights. The calculation of the cold dark matter power spectrum (Peebles, 1982; Blumenthal et al., 1984), Press & Schechter theory (Press and Schechter, 1974) and White & Rees' galaxy formation model (White and Rees, 1978)

are a few examples. However, the limits of purely analytical methods were clearly revealed as the improved datasets showed the contradictions at small scales. With the continuing advances in numerical methods and computational technologies, the structure formation and galaxy formation theories are going to be mainly conducted by numerical simulation in the future.

In this section, we will focus on dark matter only numerical simulations, especially in ΛCDM cosmology, which are well developed and have been pushed to extremely high resolution. Its only limitation is the computational resources available, in contrast to full hydrodynamic simulations which are less well-developed and limited to lower resolution. Further reading can be found in Kuhlen et al. (2012b) which reviewed numerical simulations up to recent time.

In dark matter only simulations, the density field is sampled with N particles thus "N-body" techniques can be applied to the problem. The dark matter particles are located in a coordinate system which is co-moving with the expansion of the Universe. The initial condition, which gives initial positions and velocities to all dark matter particles, is often statistically generated based on the estimated evolution of dark matter density and clustering (e.g. Klypin and Shandarin, 1983; Efstathiou et al., 1985; Crocce et al., 2006). Reed et al. (2013) claims that initial redshift should have the expansion factor $\gtrsim 10$ of that of the interested redshift, to avoid numerical artefacts.

Direct calculation of an N-body system is a $\mathcal{O}(N^2)$ problem. Techniques like the tree method (Barnes and Hut, 1986) and the adaptive particle-mesh (PM) method (e.g. ART and RAMSES) were invented to reduce the computational complexity to $\mathcal{O}(N \log N)$. An extreme method like the Fast Multipole Method (FFM) reduces the problem to $\mathcal{O}(N)$. A hybrid algorithm such as hybrid Tree-PM (e.g. GADGET; Springel et al., 2001b) is also widely used in modern cosmological simulations.

Dark matter only N-body simulations, from a few Mpc to near horizon scale (~ 20 Gpc) with more than half a trillion particles, have been performed to serve different researches in cosmology and astrophysics. The simulations can be classified by the boundary of simulation into 2 categories; full-box and zoom-in simulations. Full-box simulations resolve the entire periodic domain with a single particle mass and force resolution. They are typically used to represent a relatively large portion of the Univerese, with the box size up to a few tens of Gpc. Statistical studies of halos and large scale structure are typically the purpose of this type of simulation. On the other hand, zoom simulations are used to represent a relatively small fraction of the Universe. By concentrating the computational

power to an interested region, the highest possible resolution is used to resolve its infrastructure. The initial conditions for zoom simulations are usually nested, with multiple levels of resolution. The majority of the domain is sampled with the coarsest resolution while a small region containing the object of interest will be resolved with much higher resolution.

1.3 Structure finder

The data produced by N-body simulations is only the first step in the process to study structure and galaxy formation. The information of the particles still requires some post-processing so that their distribution can be compared to the real Universe. These analysis tools, called *halo finders*, process N-body data to find locally over-dense (either in configuration or phase-space) gravitationally bound systems, i.e. the dark matter halos which we currently believe to host galaxies (Knebe et al., 2013b).

We recommend readers to Knebe et al. (2013b) for further information and discussions. Here, we will provide only a brief summary of the halo finding algorithms and their development.

The first generation of halo finders; the spherical overdensity method mentioned by Press and Schechter (1974) (hereafter, SO) as well as the friend-of-friends algorithms introduced in astrophysical simulation by Davis et al. (1985a) (hereafter, FoF), remained standard techniques for several years, until the new methods were introduced in the 1990s. In the early time of the 21th century, halo finders became more sophisticate and were able to detect substructures; e.g. SUBFIND (Springel et al., 2001a); AHF (Knollmann and Knebe, 2009), which enable the ability to treat mergers from N-body simulations properly for the modelling of galaxy formation (Springel et al., 2001a). Before that, the approach to treat mergers were usually based on the extension of Press-Schechter formalism (Press and Schechter, 1974; Bond et al., 1991), and a Monte Carlo realisation of the hierarchical clustering process is used to statistically follow the collapse and merger history of dark matter haloes (Cole, 1991; Kauffmann et al., 1993; Baugh et al., 1996).

Even a large variety of assumptions and algorithms being used in different halo finding algorithms, we can conclude the general method of halo finding into these following steps.

- **Halo location identification and particle collection:** Most of halo finders need to identify the locations of halo candidates, either from density peaks, gravitational field or any unique criteria, before the particle collection process. However, some

algorithms such as FoF skips this step and collect particles to be the candidates for bound structures.

- **Particle unbinding:** The particles which are not gravitationally bound to the halo candidates are removed. Without this step, some halo properties such as velocity dispersion and spins can not be estimated correctly (Onions et al., 2012, 2013).

- **Boundary finding:** When a set of gravitationally bound particles are collected, the edge of the halo must be specified. Various algorithms are used in different halo finding methods. This will contribute to the properties such as the mass of haloes.

As the resolutions of N-body simulations improved over the years, the inconsistency of halo catalogues between snapshots had been revealed. The disappearance of haloes in high density regions, for example, "passing through" halo cases (see Figure 7.2) which are caused by halo finders unable to resolve multiple haloes around the density peak, will cause the overestimation of mergers. This resolution problem is also the cause of the "flipping" problem where a pair of halos can be switched between being host and satellite back and fourth. This leads to a very high fluctuation in halo mass. Therefore, some halo finders deploy the temporal information to track dark matter particles in order to reduce such problems (Han et al., 2012a). ROCKSTAR (Behroozi et al., 2013a) deploy the phase-space FoF algorithm with some additional temporal information for better identification in merging scenarios.

1.4 Merger trees

In the study of the formation and evolution of galaxies from numerical simulations, ones need to know the evolution of dark matter haloes which we assume to host galaxies. In ΛCDM, dark matter particles are only attracted to others by the gravitational force. Hence, dark matter structure grows hierarchically, with small objects collapsing under their self-gravity first and merging in a continuous hierarchy to form larger and more massive objects. By using this condition, the computational expense used to track the evolution of halos between snapshots reduces significantly and make it easier to compute in early 1990s to early 21th century (Lacey and Cole, 1993; Roukema et al., 1997; Springel et al., 2001a). This assumption implies that a halo can have only one descendant (please see the terminology used in this thesis in Section 6.2). In graph theory, a tree is an undirected simple graph which is connected and has no cycles. Therefore, we can establish

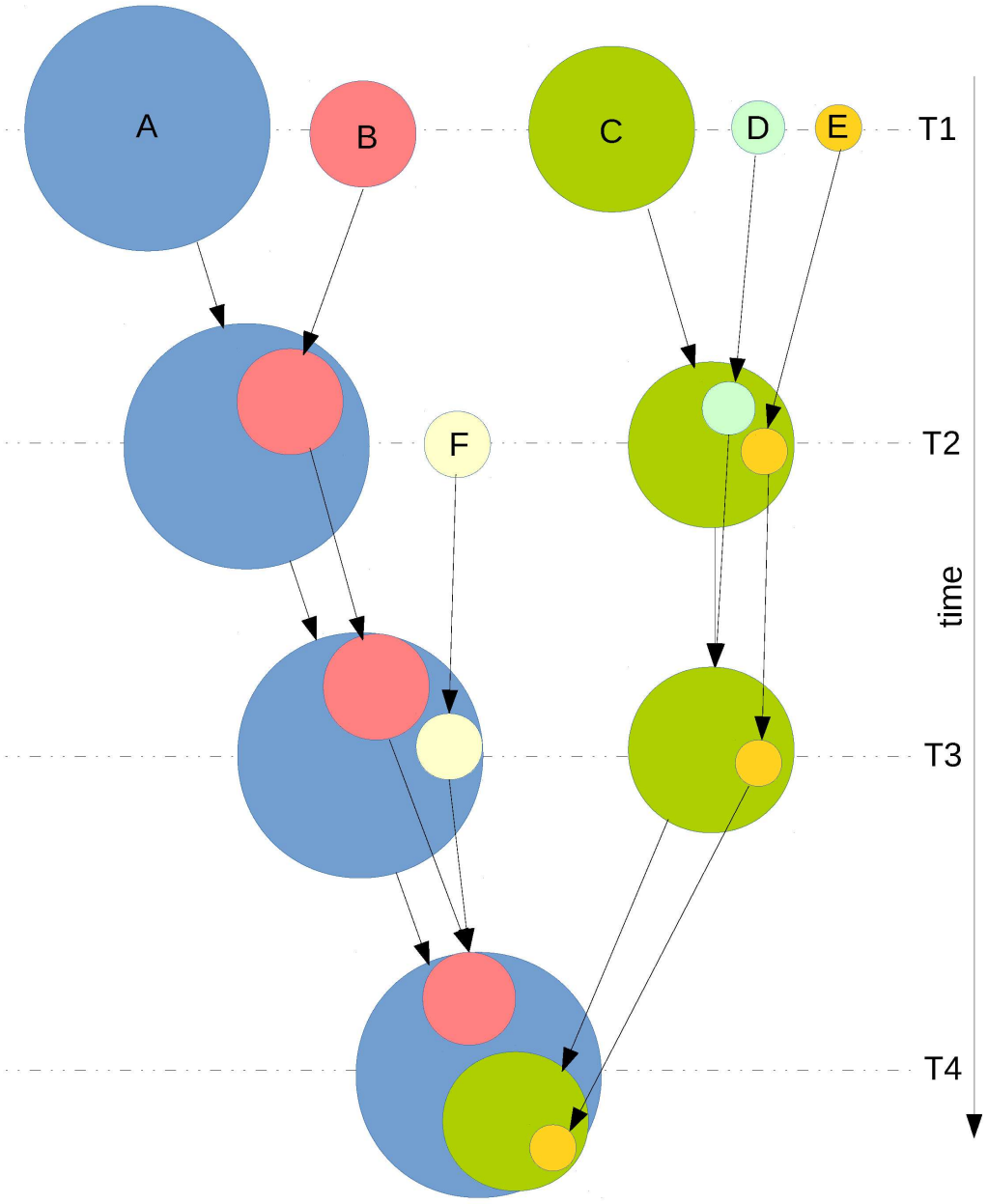

Figure 1.2: Schematic representation of the merger tree of a dark matter halo with its substructure halos along discrete time-steps.

a mathematical model by considering the relationship progenitor-descendant of haloes. We call a graph containing halos with such connection a *merger tree*.

To study the formation and evolution of galaxies, the merger trees of substructures or subhaloes are normally used instead of the merger trees of top-level haloes in order to correctly follow the baryons during the merging process. Figure 1.2 shows a simple schematic representation of the merger tree of a dark matter halo with its substructure halos along discrete time-steps. A,B,C,D,E and F are top-level halos. In the transition between T1 and T2, halo A captures halo B as its satellite, while halo C captures halo D and halo E. In the transition between T2 and T3, a newly form halo F is captured by halo A as its satellite; halo D, which is already a substructure of halo C, completely merges into halo C. In the transition between T3 and T4, halo C becomes a satellite of halo A together with halo E, which is its substructure. Halo B and and halo E which both are substructures of halo A merge together. In this case, halo B in T3 is the main progenitor of halo B in T4. Please note that the halo system shows in this figure is based on a multi-level halo scheme which allows a substructure to have substructures. The outputs from some halo finders might not allow the subhalo-of-subhalo situation hence the situation like halo E is a substructure of halo C, which is a substructure of halo A, cannot happen.

1.5 The semi-analytic approach

To study the formation and evolution of galaxies from a self-consistent cosmological model, one may directly account for the baryonic component (gas, stars, black holes, etc.) in cosmological simulation which include hydrodynamics and gravity which we call this kind of simulations *hydrodynamical simulations*. Alternatively, we can go through a procedure called semi-analytic modelling (SAM). Semi-analytic models use the distribution of dark matter haloes and their merger histories, either from cosmological simulations or the extended Press-Schechter method, together with analytical prescriptions to estimate the distribution of the physical properties of galaxies.

The great challenge for hydrodynamic simulation is the amount of computational resources makes it impractical to simulate a large enough volume necessary for galaxy surveys. Furthermore, most of the modelling in sub-grid physics in hydrodynamical simulations relies on analytical schemes similar to those used in SAMs. Therfore, it is much less costly to apply these scheme in post-processing to the simulation which is easier to find the best fit parameters the analytical sub-grid models.

The principle of semi-analytic modelling was originally introduced by White and Rees

(1978), which treats galaxy formation as a two stage process. Dark matter haloes form in a structure collapse, evolving by means of gravitational instabilities, while galaxies grow embedded in these structures, through processes that are often dissipative and non-linear.

The first self-consistent models based on SAM methodology were developed by Cole (1991), Lacey and Silk (1991) and White and Frenk (1991). During that time, most of the modern physics recipes were already presented, e.g. the collapse of gas into an hot phase and subsequent cooling into cold disc, the formation of stars from the cold disc, supernova feedback, chemical enrichment to account the effect from supernovae on the properties of the hot gas and stellar population synthesis model to estimate photon energy from star formation histories. However, Press-Schechter formalism (Press and Schechter, 1974; Bond et al., 1991), and a Monte Carlo realisation of the hierarchical clustering process are widely used to statistically follow the collapse and merger history of dark matter haloes until the end the 20th century.

Modern SAMs have shifted the focus towards the utilisation of merger trees directly generated from N-body simulations instead of the using analytic formalisms as before. They can reliably capture non-linear structure formation which allow us to study the physics of galaxy formation in smaller scale.

We will give a very brief summary of how SAMs works (see the details of a particular SAM, L-Galaxies model, in Section 4.3). The main physical models used in SAMs are briefly summarised as follows (White and Frenk, 1991; Kauffmann et al., 1993; Cole et al., 1994; Somerville and Primack, 1999; Cole et al., 2000; Somerville et al., 2001; Croton et al., 2006; De Lucia et al., 2006; Cattaneo et al., 2007; Somerville et al., 2008):

- A certain amount of gas from the IGM is shock-heated as it joins a halo.
- The hot gas cools via atomic line transitions and other mechanisms.
- The cooled gas collapses to form a galactic disc.
- Stars are formed inside the galactic disk.
- Radiation from massive stars and supernovae reheat the gas inside the halo.
- Mergers trigger bursts of star formation. Major mergers transform discs into spheroids/ellipsoids.
- Other models, such as the stellar population synthesis and dust extinction model, are used to estimate observational properties.

The complexity due to the large number of physical models used in SAMs requires efficient calibration tools to adjust the free parameters according to observational constraints. Without an automatic tool, e.g. Monte Carlo Markov Chain (MCMC) parameter estimation (Trotta, 2008; Henriques et al., 2009), to sample the coordinates in high-dimensional parameter spaces, it is nearly impractical to produce a model with robust prediction power in modern astrophysics.

An example of successful SAMs is the semi-analytical models based on the Millennium Simulation (Springel et al., 2005b) and Millennium Simulation II (Boylan-Kolchin et al., 2009), which are able to replicate many statistical results from observed properties of galaxies (Springel et al., 2005b; Crocce et al., 2006; De Lucia and Blaizot, 2007; Guo et al., 2011, 2013; Henriques et al., 2015).

Chapter 2

The Epoch of Reionisation

2.1 The first light

About 400,000 years after the Big Bang, the density and the temperature of the Universe decreased. When the Universe was cool enough, it allowed ions to recombine into neutral hydrogen and helium atoms. After that, photons decoupled from baryons and radiated through the transparent Universe, leaving the Cosmic Microwave Background (CMB) radiation which is able to observed today. This period is known as the *dark ages*.

About 400 million years later, the first galaxies formed and started to emit ionising photons. This brought the end of the *dark ages* and the Epoch of Reionsation (EoR) began. The intergalactic medium (IGM) was initially ionised around the first objects. After that, the patchy regions of neutral (H I) and ionised hydrogen (H II) evolved through the ionisation and recombination. When there were enough number of ionising photons from UV emitting sources, the temperature of the gas in the Universe increased drastically. The ionised gas regions also evolved and eventually filled out the entire Universe (Barkana and Loeb, 2001; Loeb and Barkana, 2001; Bromm and Larson, 2004; Ciardi and Ferrara, 2005; Choudhury and Ferrara, 2006; Furlanetto et al., 2006).

The current constraints suggest that the EoR roughly occurs within the redshift $6 \lesssim z \lesssim 15$. However, the detailed process involving reionisation is still uncertain. The factors which controlled the formation of the first objects and ionising photon intensity they produced is not well known. The topology of expanding ionisation bubbles in the intergalactic medium is not yet to be confirmed. In order to answer these questions and many other factors which may arise in the study of the EoR, we need the know the fundamental issues in cosmology, galaxy formation, quasars and the physics of Population III stars. There are significant theoretical and observational efforts dedicated to understand-

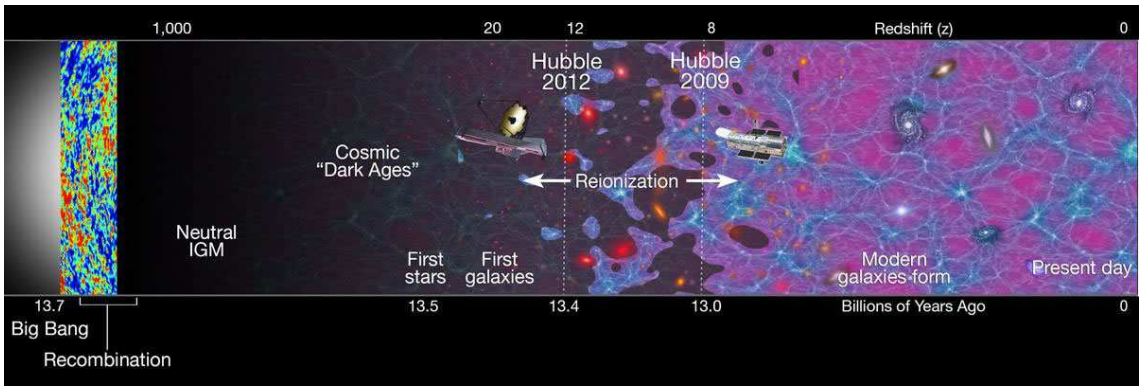

Figure 2.1: Cartoon of the transition from the neutral intergalactic medium (IGM) left after the Universe recombined at $z \sim 1100$ to the fully ionised IGM observed today. Image credit: NASA/ESA, R. Ellis (Caltech), and the UDF 2012 Team.

ing the processes in the evolution of EoR, including the effects on structure formation (see more details in Barkana and Loeb, 2001; Bromm and Larson, 2004; Ciardi and Ferrara, 2005; Choudhury and Ferrara, 2006; Furlanetto et al., 2006). However, observational support is very limited, and when available, is indirect and model dependent.

2.2 Observational evidences for the Epoch of Reionisation

2.2.1 Gunn-Peterson troughs

The intergalactic medium (IGM) can be studied through the analysis of the Lyman-α forest, which is an absorption lines seen in the spectra of background quasars or quasi-stellar objects (QSOs). Gunn and Peterson (1965) predicted that the neutral hydrogen between an observer and a quasar, would suppress the radiation from the quasar. The Gunn and Peterson optical depth in a uniformly distributed IGM is given by:

$$\tau_{\rm GP}(z) = \frac{1.8 \times 10^5}{h} \Omega^{1/2} \left(\frac{\Omega_b h^2}{0.02}\right) \left(\frac{1+z}{7}\right)^{3/2} \left(\frac{n_{\rm HI}}{n_{\rm H}}\right). \tag{2.1}$$

We can see that even only a small neutral fraction can make $\tau_{\rm GP} >> 1$ and produce an almost-complete absorption of the UV resonance line frequency of 1215.67Å(rest-frame). We call this phenomenon a Gunn-Peterson trough. It was first observed in the spectrum of a quasar at $z = 6.28$ with the Sloan Digital Sky Survey (SDSS) in 2001 (Becker et al., 2001). Several samples of high redshift quasars have been identified ever since (Djorgovski et al., 2001; Fan et al., 2002, 2006; White et al., 2003). The results show that the volume-averaged neutral hydrogen fraction increases from 10^{-5} at $z \sim 3$ to greater than 10^{-3} at

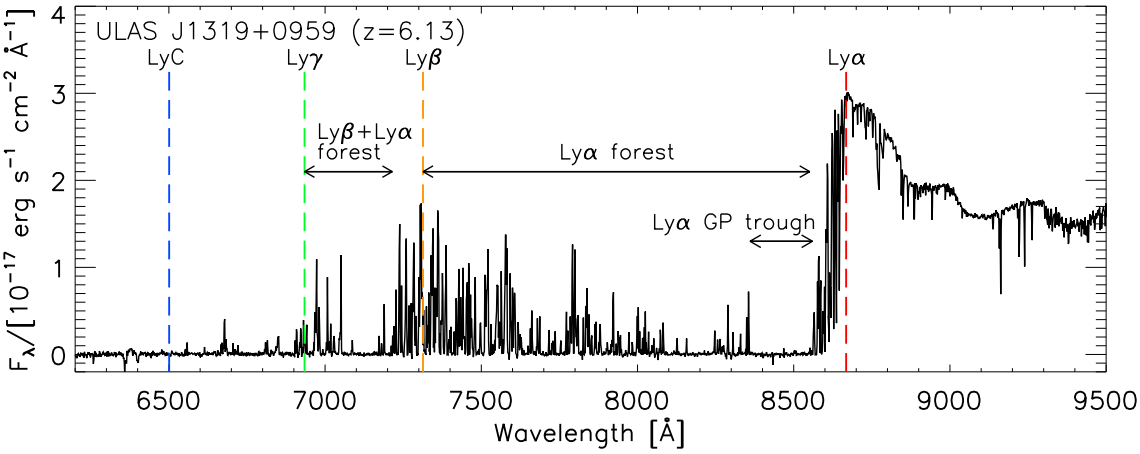

Figure 2.2: A high signal-to-noise spectrum of the quasar ULAS J1319+0959 at $z = 6.13$ from Becker et al. (2015a), obtained with the X-Shooter spectrograph on the Very Large Telescope (VLT), shows the observed Gunn-Peterson troughs.

$z \sim 6$ (Fan et al., 2002). This discovery combined with the absence of Gunn-Peterson trough in the spectra of the quasars at $z < 6$ weakly constrain the end of the EoR.

2.2.2 The Thomson scattering optical depth for the cosmic microwave background (CMB) radiation

Another important constraint of reionisation is the Thomson scattering of CMB photons. This topic has been discussed in detail by many authors (Peebles and Yu, 1970; Sunyaev and Zeldovich, 1972; Bond and Efstathiou, 1984; Ma and Bertschinger, 1995; Hu and White, 1997; Aghanim et al., 2008). The CMB photons were scattered by the free electrons (i.e. from ionised hydrogen atoms) and caused the suppression in the temperature fluctuation by $1 - e^{-\tau}$(Zaldarriaga et al., 2004), where τ is the optical depth. The observed value for the optical depth is 0.067 ± 0.016 (*Planck* TT + lensing + BAO)(Planck Collaboration et al., 2015). The optical depth along a line of sight can be expressed in terms of the integral

$$\tau_e = \int_0^{z_{dec}} \sigma_T n_e \frac{cH_0^{-1} dz}{(1+z)\sqrt{\Omega_m (1+z)^3 + \Omega_\Lambda}}, \tag{2.2}$$

where z_{dec} is the decoupling redshift, σ_T is the Thomson cross section and n_e is the electron density. This relationship also works for the mean electron density of the Universe; hence it could be turned into a constraint on the reionisation history of the Universe.

2.2.3 21-cm hyperfine hydrogen line

The redshifted 21-cm line has been successfully used to probe the dynamic of the neutral hydrogen in the Milky Way and other galaxies (Warmels, 1988; Dickey and Lockman, 1990; Kalberla and Kerp, 2009; Cortese et al., 2010). It is potentially a mean to study the *dark ages* and the EoR. This topic has been discussed and reviewed in detail by many authors (Furlanetto et al., 2006; Pritchard and Loeb, 2012; Furlanetto, 2015). Only a brief review will be discussed here. The 21-cm line is from the transition between the two hyperfine levels of the hydrogen $1s$ ground state with an energy difference of 5.87433 μeV. This corresponds to a 21.106 cm (1420.405 MHz) photon. The transition is highly forbidden with a very small rate of 2.85×10^{-15} s^{-1}.

The redshifted 21-cm is defined as the differential brightness temperature, δT_b, as

$$\delta T_b = T_S(1 - e^{-\tau_\nu}) + T_B e^{-\tau_\nu} - T_B, \tag{2.3}$$

where T_B is the brightness temperature of the background source. The optical depth $\tau_\nu \equiv \int \alpha_\nu ds$ is the integral of the absorption coefficient along a line of sight, and s is the proper distance (Field, 1959). For a cloud of hydrogen, $\tau_\nu \sim \sigma_{01} \left(\frac{h\nu}{k_B T_S}\right) \left(\frac{N_{\rm HI}}{4}\right) \sigma(\nu)$ where $\sigma_{01} \equiv \frac{3c^2 A_{01}}{8\pi\nu^2}$, $A_{01} = 2.85 \times 10^{-15}$ is the spontaneous emission coefficient of the 21-cm transition, $N_{\rm HI}$ is the column density of H I , and $\sigma(\nu)$ is the normalised line profile.

The spin temperature T_S, which is determined by the absorption and emission of the 21-cm photons with the temperature $T_{\rm CMB}$, can be expressed as

$$T_S = \frac{T_{\rm CMB} + x_c T_K + x_\alpha T_c}{1 + x_c + x_\alpha}, \tag{2.4}$$

where x_c and x_α are the normalised probabilities of the collisions and Lyman-α scattering respectively. T_K is the gas kinetic temperature and T_α is the temperature due to the scattering of Lyman-α photons.

When the source of Lyman-α photons has formed, the neutral hydrogen in the IGM absorbs the emitted 10.2 eV photons and returns to one of the hyperfine spin states. The coupling of the neutral hydrogen with the Lyman-α radiation is called *Wouthuysen-Field* effect (Wouthuysen, 1952; Field, 1958). This phenomenon has the effects on the 21-cm signal variation at $\sim 10\%$ level.

By studying 21-cm line emission and absorption, it will be possible to learn more about the formation of the first structures. There are many projects dedicated to measure this faint signal. In the short term, there are the Precision Array to Probe the Epoch of Reionization (PAPER) (Backer et al., 2007), the Low-Frequency ARray (LOFAR) (Brentjens

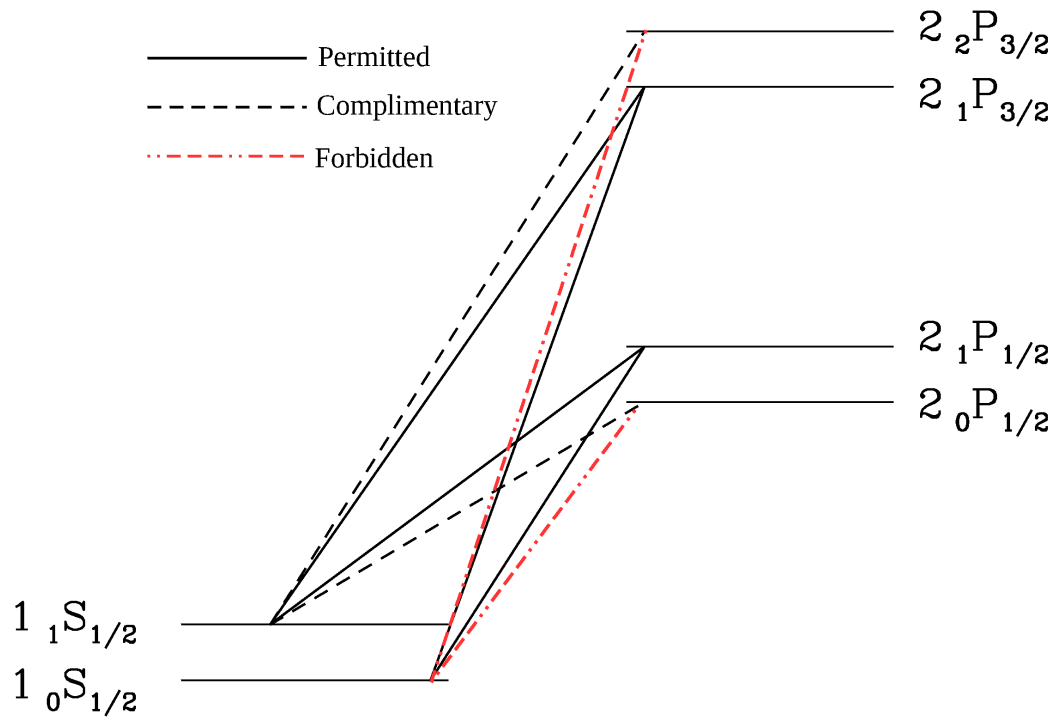

Figure 2.3: The hyperfine structure of the hydrogen atom and the transitions relevant for the Wouthuysen-Field effect. Black solid line transitions allow spin flips, while black dashed transitions are allowed but do not contribute to spin flips, and the red dot-dot-dashed transition lines are strictly forbidden (Pritchard and Loeb, 2012).

et al., 2011), the Murchison Widefield Array (MWA) (Tingay et al., 2013), and the Giant Metrewave Radio Telescope (GMRT) (Pen et al., 2009). In the longer term, there is the Square Kilometer Array (SKA) (Cordes, 2005) and its pathfinders. Although there is no direct 21-cm measurement at the moment, Parsons et al. (2014) used 3-months data using 32 antennas of PAPER to place the upper limit of the power spectrum of 21-cm emission within an order of magnitude of the brighter predictions of the expected 21 cm EoR signal level. Dillon et al. (2014) also used the observations from the 32-tile prototype of the MWA to establish upper limits on the power spectrum from $z = 6.2$ to $z = 11.7$.

2.2.4 Other observational probes

There are other observations that may provide evidence for reionisation, for example, the direct studies of the high redshift ($z \lesssim 6$) IGM such as the IGM temperature evolution and the number of ionising photons per baryon. The kinetic Sunyaev-Zel'dovich effect

on the CMB caused by the reionisation is also a probe to study the reionisation history, see Park et al. (2013) for details. Other probes include cosmic infrared and soft x-ray background (Dijkstra et al., 2004a), Lyman-α emitters (Ouchi et al., 2009), high redshift QSOs (Mortlock et al., 2011) and gamma-ray burst (GRBs) (Bromm and Loeb, 2006). However, these probes currently provide very limited constraints on the EoR. See Zaroubi (2013) for the detailed discussion of the probes.

2.3 Sources of reionising photons

To ionise a neutral hydrogen atom, the energy larger than 13.6 eV is needed. This corresponds to photons with a maximum wavelength of 912 Å. Therefore, the primary candidates are all sources which produce sufficient amount of energy in UV and above. Other factors such as the photon density and the lifetime of the sources must be considered as well. With these constraints, we can expect that the first generation stars and galaxies, including quasars were the main source of energy (Madau et al., 1999; Barkana and Loeb, 2001).

2.3.1 Quasars and Active Galactic Nuclei (AGN)

Quasars, or quasi-stellar objects (QSOs), and AGNs are effective emitters of UV photons, as the escape fraction of photons is assumed to be of the order of unity. However, unless there is a large population of those with very low luminosity, the emissivity of QSOs does not meet the requirement for reionisation (Bolton and Haehnelt, 2007; Meiksin, 2005; Srbinovsky and Wyithe, 2007). Their observed population could not produce enough UV radiation to reionise the IGM at redshift $z \sim 3$ and their density declines exponentially towards high redshifts (Shapiro and Giroux, 1987; Madau et al., 1999; Giroux and Shapiro, 1996). Another important point about QSOs is their short lifetime, which might affect the growth of their ionized regions. Moreover, unless they are extremely X-ray quiet, their contribution to the reionisation history should be $< 30\%$ with respect to star-forming galaxies (Zaldarriaga et al., 2003). They are instead the main sources for Helium reionisation. QSOs likely fully ionised helium by $z \sim 3$, and possibly as early as $z \sim 5$ if they have hard spectra extending down to the He II photoelectric (Meiksin, 2005). However, there is still uncertainty on the estimate of the total UV photon emissivity from star-forming galaxies at high redshifts as there are still unknown factor like the stars formation rate, the clumping factor and the UV escape fraction (see Section 2.3.2).

2.3.2 Star-forming galaxies

Hot-young stars in star-forming galaxies are the source of Lyman continuum photons. Due to the decrease of QSO population in high redshifts, even their spectra are too soft to produce enough He II -ionising photons, they are the primary candidate sources for hydrogen-ionising photons (Bouwens et al., 2012, 2015a; Robertson et al., 2015). However there is still uncertainty in estimating the total UV photons emitted from these galaxies, as such an estimate requires several theoretical assumptions regarding the Initial Mass Function (IMF), populations, ages, and spectra of the massive stars, which dominate the radiation at early times. A large uncertainty is the star-formation rate at high redshifts: observational estimates are plagued by sample variance at $z \sim 6$, due to the small volume of deep surveys, as well as extrapolations toward faint luminosity objects. Moreover, another major uncertainty is the escape fraction of the ionising radiation, as this must multiply the emissivity based on galaxy counts and finally governs the contribution of the star-forming galaxies to the UV background which is still uncertain due to several factors. There are few direct measurements of the escape fraction, and virtually all are upper limits. Putman et al. (2003) infer a global escape fraction from the Milky Way of 12% from Hα spectroscopic measurements of nearby high velocity clouds. Based on a detection of Lyman continuum flux from FUSE observations of a local starburst galaxy, Bergvall et al. (2006) suggest an escape fraction of $0.04 - 0.1$, subject to spectral modeling uncertainties. Fernández-Soto et al. (2003) place an upper limit of 4% for galaxies in the redshift range $1.9 < z < 3.5$. Similar upper limits based on FUSE observations were obtained by Deharveng et al. (2001) and Heckman et al. (2001). Shapley et al. (2006) report the detection of UV radiation shortward of the Lyman edge in a sample of Lyman Break Galaxies at $z \sim 3$ with a corresponding escape fraction of 14%, although the estimate is sensitive to the assumed calibrating population synthesis models and corrections for the IGM optical depth, and it may also apply only to the more luminous galaxies. Inoue et al. (2006) estimate the the escape fraction to be increasing with redshift, from a value less than 0.01 at $z \lesssim 1$ to about 0.1 at $z \lesssim 4$.

On top of the ionising radiation escape fraction and UV-to-Lyman limit conversion factor, the effects of internal galactic extinction, uncertainty in the amount of intergalactic extinction by the IGM and the minimum source luminosity on the UV luminosity density and its evolution are source of uncertainties in the net contribution of the galaxies. All errors combined together still leave a range of uncertainties of about a factor of 4 for the contribution of the galaxies to HI, excluding the uncertainty of the escape fraction (Ferrara

and Pandolfi, 2014).

2.3.3 Population III stars

Population III stars are the stars which are made of only hydrogen and helium. There is an evidence for such stars existed at $z = 6.6$ (Sobral et al., 2015). There are also indirect observations using gravitational lensing (Fosbury et al., 2003). These Population III stars are more efficient and effective than Population II stars since they emit more reionising photons (Tumlinson et al., 2002). They are currently considered to be the energy sources to initiate the reionisation of the Universe (Alvarez et al., 2006). However, the contribution of these massive metal-free stars is uncertain, as producing a number of photons able to ionize the IGM would lead to an overproduction of metals. This implies that the transition to metal-enriched Pop II stars would occur earlier. This problem will be solved if Population III stars collapse to form black holes, then the ionising radiation from the accretion on SN-induced seed black holes might be more significant than the primary emission (Ricotti and Ostriker, 2004). Another suggestion is to differentiate very-massive metal free stars and metal-free stars with masses less than $100M_{\odot}$, as in this way the formation of lower mass Pop III stars would stop at lower redshifts due to metal enrichment and feedback, and therefore their contribution would be more prominent (Schneider et al., 2006).

2.3.4 Other possible sources

Other possible sources for UV photons are: mini-quasars powered by intermediate mass black holes (Madau et al., 2004); annihilations from Dark Matter particles like decaying sterile neutrinos (Hansen and Haiman, 2004; Mapelli and Ferrara, 2005); X-ray photons from binary stars in early mini-quasars, but also from thermal emission from the supernova remnants, or Compton up-scattering of CMB photons by relativistic electrons produced by supernova explosions (Venkatesan et al., 2001; Madau et al., 2004; Ricotti et al., 2005); enhanced structure formation from a non-scale-free isocurvature power spectrum (Sugiyama et al., 2004); and non-Gaussian density fluctuations (Chen et al., 2003).

2.4 Reionisation of hydrogen in the IGM

There are several stages involving the reionsation of hydrogen in the IGM. We will give a brief summary of this topic as an introduction. Further reading can be found in Gnedin (2000); Loeb and Barkana (2001); Barkana and Loeb (2001); Meiksin (2009); Ferrara and Pandolfi (2014); Robertson et al. (2010).

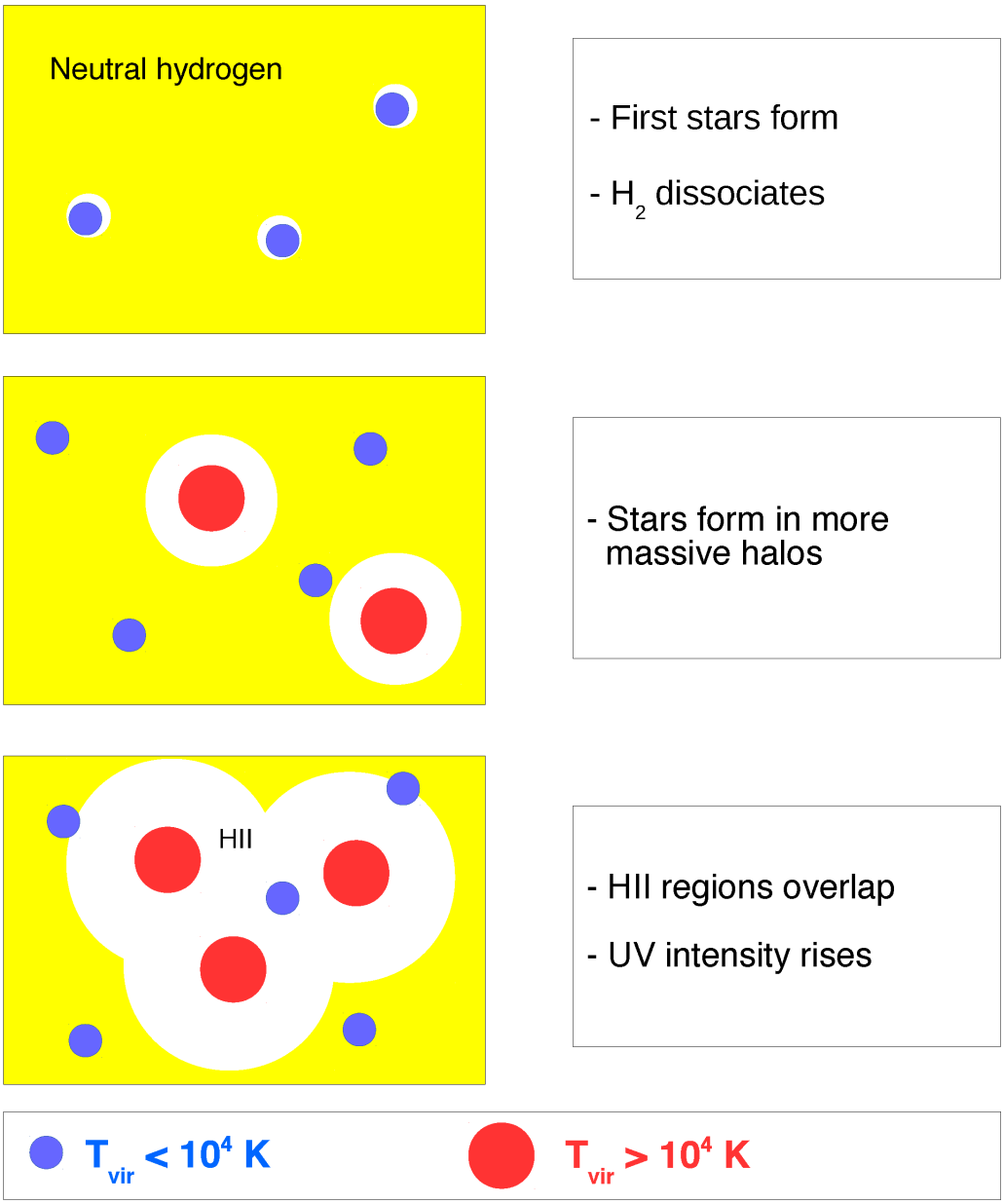

Figure 2.4: Processes in the reionisation of hydrogen in the intergalactic medium.

Initially, the ionising photons from an individual ionising source begin to ionise its surrounding IGM. The first generation galaxies usually form in the most massive haloes at high redshift, which are biased and clustered around the high density peaks. Due to the dense baryon contents, which is characterised by a high recombination rate, the ionising photons emitted from the galaxies will have a short mean free path which make it difficult to leave the high density peaks. When they are able to escape, the ionisation fronts propagate more easily into the lower density regions, leaving pockets of neutral, high density gas behind.

Once the neighbouring H II bubbles begin to overlap, any point in the merged regions will be exposed to ionising photons from the sources in both former regions. Therefore, the ionising photon density increases drastically and allows the combined regions extend into high density gas in the neutral pockets, which were left behind due to the low ionising intensity. As more H II bubbles overlap, combined regions are exposed to more ionising sources, which accelerates the growth of ionised regions. At the end of this stage, the IGM is highly ionised except the gas located inside self-shielded high density clouds.

Some gas can remain neutral within high-density structures, e.g. Lyman Limit systems and damped Lyman-α systems, which can be observed at lower redshifts. The high-density regions are progressively ionised as galaxies continue to form, which increase the ionising photon abundance over time. When the ionising intensity is sufficient, the entire Universe is highly ionised and exposed to ionisation radiation, which marks the end of the EoR.

2.5 Photo-evaporation of the gas in minihaloes during the EoR

Virialised minihaloes are small haloes with virial temperatures, $T_{\rm vir}$, below the limit of atomic cooling transtion temperature, 10^4 K. These minihaloes are in the mass range between the Jeans mass, minimum mass of a dark matter halo to retain baryons from escaping to the IGM, i.e. $M_{\rm J} \sim 10^4\ M_{\odot}$, and the mass for which $T_{\rm vir} = 10^4$ K which corresponding to $\sim 10^8\ M_{\odot}$. Their baryonic content is highly neutral because the temperatures are too low for collisional ionisation to be effective Shapiro et al. (2003); Iliev et al. (2002, 2005).

As the gas in minihaloes gets exposed to ionising radiation, either from first generation stars inside the minihaloes or the external UV background, it could be heated up to ∼1-2×10^4 K. A major fraction of the collapsed gas, which had been bound to the minihaloes

with a virial temperature less than $\sim 10^4$ K, could be evaporated back into the IGM, due to the lack off efficient cooling. These gas-poor minihaloes will affect the galaxy formation just after reionisation, as well as at lower redshifts (Loeb and Barkana, 2001).

Figure 2.5, (taken from Figure 2 in Okamoto et al., 2008), shows the baryon fraction inside halos as function of their total mass, M, for simulated haloes at redshifts $z \sim 9.3$ (left), 5 (middle), and 2 (right) for all haloes in the simulations (top row) and "isolated haloes", which are the haloes outside a radius of $6R_{\rm vir}$, where $R_{\rm vir}$ is the virial radius of the halo, from any haloes (bottom row). The cosmological baryon fraction ~ 0.18 is indicated as the dotted horizontal line. Plus signs and diamonds indicate haloes from the low resolution and high resolution simulations respectively. The plots show, just after the reionisation at $z = 5$, the baryon fraction of halos with mass less than the Jean mass of the gass at $T \sim 10^4$ K, 10^9 $M_\odot$ gradually drops down to zero at around M $\sim 10^8$ $M_\odot$. Long after the reionisation at $z \sim 2$, the gas of the haloes with mass less than 10^9 $M_\odot$, is almost gone (this will be discussed in the next section). Please note that the simulation used in Okamoto et al. (2008) assumes uniform external UV background radiation without including the effect from local photonisation.

2.6 Suppression of the formation in low mass galaxies

After reionisation, the UV photons heat up the IGM to above $\sim 1 - 2 \times 10^4$ K, the linear theory estimates the instantaneous Jeans mass to be

$$M_{\rm J} = 4.1 \times 10^9 \ {\rm M}_\odot \left(\frac{T}{10^4 \ {\rm K}}\right)^{3/2} \left(\frac{h^2\Omega_m}{0.1327}\right)^{-1/2} \left(\frac{h^2\Omega_b}{0.0216}\right)^{-3/5} \left(\frac{1+z}{10}\right)^{3/2}, \qquad (2.5)$$

which is roughly $\sim 10^9$ $M_\odot$ (e.g. Shapiro et al., 2004; Iliev et al., 2002, 2008a). However, even in a minihalo with the mass less than the instantaneous Jeans mass, a fraction of baryons from IGM still can collapse because the mass scale on which baryons successfully collapse out of the IGM along with the dark matter is determined by integrating the differential equation for perturbation growth over time for the evolving IGM (Shapiro et al., 1994; Gnedin and Hui, 1998; Gnedin, 2000). A halo collapsing inside an ionised region can only acquire enough gas to form stars if it is sufficiently massive. The minimum mass depends on the detailed gas dynamics of the process and on radiative heating and cooling. The drops in the baryon fraction in Figure 2.5 demonstrates the smooth transition between the mass which a collapsing halo retains all its gas, and halo mass which the gas does not collapse with the dark matter, not a sharp mass threshold to determine which

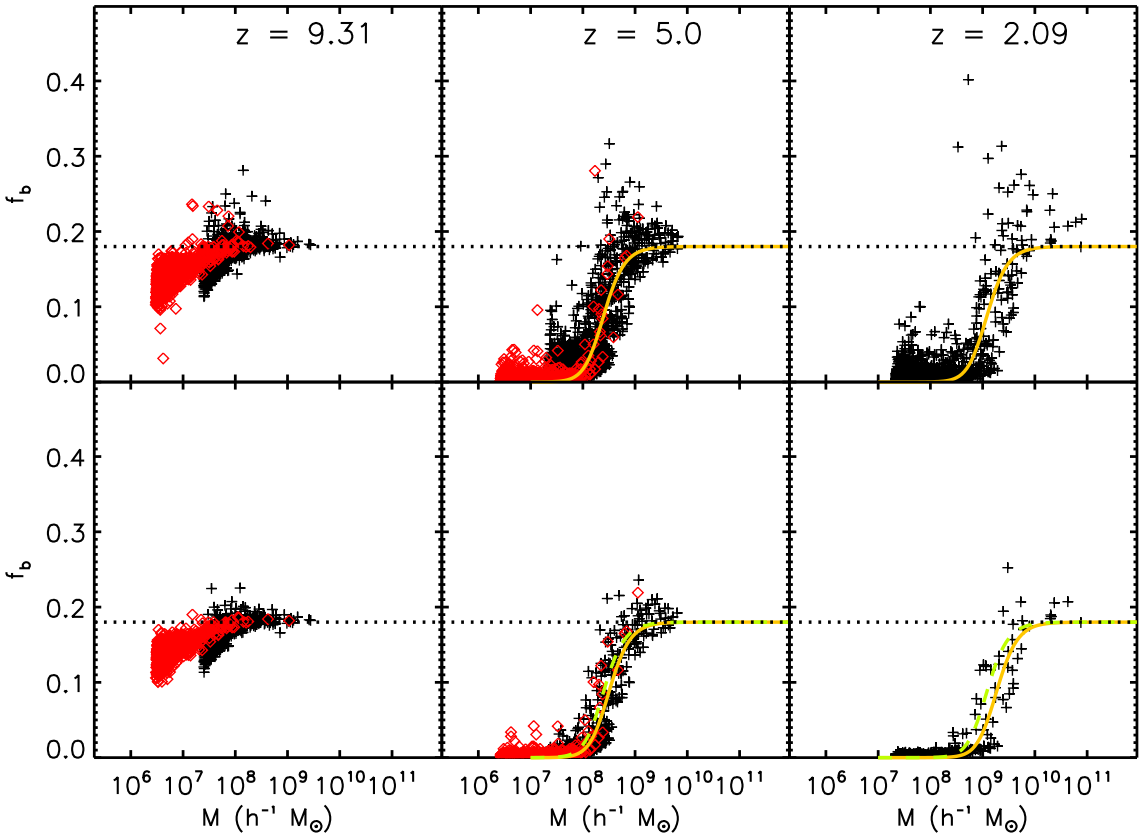

Figure 2.5: Fraction of baryonic mass as function of total mass, M, for simulated haloes at three redshifts z for all haloes (top row) and "isolated haloes", which are the haloes outside the $6R_{\rm vir}$ radius from any haloes (bottom row). The cosmological baryon fraction ~ 0.18 is indicated as the dotted horizontal line. Plus signs and diamonds indicate haloes from the low resolution and high resolution simulations respectively (Figure. 2, Okamoto et al., 2008).

halo can or cannot accumulate baryons from the IGM.

Part II

Semi-Analytic Modelling with Radiative Feedback

Chapter 3

Introduction

Understanding the physics of the IGM during the Epoch of Reionisation (EoR) is one of the most important problems in modern cosmology. It offers crucial information about the structure formation and physical processes in the early Universe. The discovery of high redshift quasars by the Sloan Digital Sky Survey (SDSS) demonstrated that reionisation was mostly completed by $z \sim 6$ (Becker et al., 2001; Fan et al., 2006) but might fully finish at $z \sim 5$ (Becker et al., 2015b). Furthermore, absorption spectra from some higher redshift quasars provide the evidence of a partially neutral IGM (e.g. Fan et al., 2006; Becker et al., 2015b), although most of the interpretations are still under debate (Lidz et al., 2006; Bolton and Haehnelt, 2007; Mesinger and Furlanetto, 2009). In order to observe the geometry of ionised and neutral patches of hydrogen, the next generation radio telescopes such as PAPER, LOFAR, MWA, and SKA are designed to be able to detect these features.

The theoretical methods to study the inter-galactic hydrogen during the EoR have been rapidly developed in the past decades. Traditional approaches adopt analytic modelling to investigate the average behaviours on the large scales but they cannot address the complex feedback processes since the calculations are linear (e.g. Furlanetto and Loeb, 2003; Wyithe and Loeb, 2004, 2013). Numerical methods, on the other hand, include non-linear recipes but require a large amount of computational time. These methods usually begin with a collisionless dark matter N-body simulation and then halo catalogues are extracted by a halo finder. Afterwards, a simple prescription is used to convert halo mass to the ionising photon emitting rate. Then, a radiative transfer method will be used to calculate the ionisation structure on the large scales (Mellema et al., 2006; Gnedin and Abel, 2001; Maselli et al., 2003; Razoumov and Cardall, 2005; Ritzerveld et al., 2003; Whalen and Norman, 2006; Rijkhorst et al., 2006). There are some methods employing N-body simulation together with analytical methods, called *semi-numerical* methods, which

are able calculate the reionisation structure in very large volume with a little computational resources compared to a radiative transfer method. (Mesinger and Furlanetto, 2007; Zahn et al., 2007; Geil and Wyithe, 2008; Lidz et al., 2009; Choudhury et al., 2009; Alvarez et al., 2009; Santos et al., 2010). However, these methods only have the ability to construct the reionisation structures and their evolution, but do not provide any information about the galaxy formation and evolution in the EoR.

In order to understand the physics of galaxy formation, many hydrodynamic simulations have been utilised to directly model the star formation in high redshift galaxies and couple with sub-grid physical recipes such as feedback and metal enrichment. These simulations are capable of producing the luminosities of the galaxies at high redshift (e.g. Finlator et al., 2011; Salvaterra et al., 2011; Jaacks et al., 2012; Genel et al., 2014; Furlong et al., 2015). However, a large amount of computational resources used by these simulations make it impractical to simulate a large enough volume necessary for galaxy survey. Another approach to model high redshift galaxies is to utilise semi-analytic models (SAMs) of galaxy formation (Lacey and Cole, 1993; Kauffmann et al., 1993). This method generally requires relatively less computational expense than those of hydrodynamic simulations of the same simulation size.

The frontier of galaxy survey has been expanded dramatically since the installation of Wide Field Camera 3 (WFC3) on the *Hubble Space Telescope (HST)* in 2009 (e.g. Oesch et al., 2010; Bouwens et al., 2010b; Bunker et al., 2010; Wilkins et al., 2010; Finkelstein et al., 2010; McLure et al., 2010; Wilkins et al., 2011a,b; Lorenzoni et al., 2011; Bouwens et al., 2011; McLure et al., 2011; Finkelstein et al., 2012a,b; Lorenzoni et al., 2013; McLure et al., 2013; Duncan et al., 2014). The rest frame ultraviolet (UV) luminosity function (LF), which is an important probe to constraint the star formation rates of the galaxies in the early Universe, of almost 1000 galaxies at $z \geq 7$ can be observed (Bouwens et al., 2015b; Finkelstein et al., 2015). Furthermore, the next generation space observatory, the *James Webb Space Telescope (JWST)* will be launched in a few years. However, the ability to observe the faintest galaxies, which are expected to play the crucial role in driving the reionisation of the neutral hydrogen in the inter galactic medium (IGM) after the dark ages, will not be implemented in any observation in the near future (Robertson et al., 2013; Duffy et al., 2014). It has been suggested that UV radiation during the EoR is responsible for suppressing the star formation of these faintest galaxies. The suppression is believed to be caused by 2 factors. One is from the photo-evaporation of the gas in minihaloes, which are the small objects with virial temperature below the atomic line

cooling temperature of 10^4 K (Barkana and Loeb, 1999; Shapiro et al., 2004; Iliev et al., 2005). The other factor is the increase in pressure of the intergalactic gas due to the photoheating from galaxies, which will suppress the baryonic infall process in the galaxy formation and evolution (Shapiro et al., 1994; Gnedin, 2000; Hoeft et al., 2006).

Accounting for the suppression in small galaxies provides a solution to the "missing satellites problem" (Klypin et al., 1999; Moore et al., 1999). The common approach is to impose a simple parametrised model to approximate the average ionising background as a function of redshift (Gnedin, 2000) in both hydrodynamic simulation Haardt and Madau (1996, 2012); Okamoto et al. (2008); Pawlik et al. (2013); Wise et al. (2014); Jeon et al. (2014); Rosdahl et al. (2015); Aubert et al. (2015) and semi-analytic modelling of galaxy formation (Croton et al., 2006; Bower et al., 2006; De Lucia and Blaizot, 2007; Guo et al., 2011; Henriques et al., 2015). There are hydrodynamical simulations of galaxy formation with coupled radiative transfer have been used to address the effects of reionisation on galaxy formation self-consistently recently (So et al., 2014; Norman et al., 2015). Recent works such as Gnedin (2014) and Pawlik et al. (2015) managed to run this state-of-art radiative hydrodynamic simulations with moderated box sizes and resolutions. Ocvirk et al. (2016) also used a radiative hydrodynamic simulation of the local Universe to study the suppression of the star formation by photo-ionisation heating. However, these simulations came with a tremendous amount of computational expense, much greater than those of normal hydrodynamic simulations. Furthermore, all simulations require calibrations of sub-grid models. This leads to even more difficult computational tasks to overcome.

While the pure numerical methods are computationally impracticable, the semi-analytic modelling of galaxy formation is a feasible approach to address this problem. It requires much less computational resources but gives a robust prediction of high redshift galaxies (Benson et al., 2006; Lacey et al., 2011; Raičević et al., 2011; Zhou et al., 2013; Clay et al., 2015). By fully combining a SAM with a reliable recipe of reionisation, one can self-consistently model the suppression of star formation from reionisation and the effects of the reionising photon emission from galaxies on the inter-galactic medium. Recent works by *DRAGON* project use this method and produced good results which agree with observations (Graziani et al., 2015; Mutch et al., 2015; **?**).

In this work, we used the very well-tested Munich semi-analytic modelling of galaxy formation, L-Galaxies (Kauffmann, 1996; De Lucia and Blaizot, 2007; Guo et al., 2011, 2013; Henriques et al., 2015), coupled with the semi-numerical method of reionisation simulation described in Majumdar et al. (2014), on a very high resolution N-body simulation

described in Section 4.1. The combination allows more capability in modelling the effect from reionisation. We then use the newly created pipeline to create galaxy catalogues and H II density fields which are investigated in Chapter 4. The framework used in this work was developed to be flexible ensuring the compatibility with the new public version of L-Galaxies and a radiative transfer simulations, e.g. C^2-ray (Mellema et al., 2006), which we will integrate into the pipeline in the near future.

Chapter 4

Simulations and Results

4.1 The N-body dark matter simulation

A very high resolution N-body simulation of the formation of high-redshift structures in Dixon et al. (2016) is used in this work to provide sufficient mass resolution to accurately capture the low-mass galaxy population driving the reionisation of the Universe at high redshift. It was generated by the cosmological N-body simulation code CubeP^{3}M (Harnois-Déraps et al., 2013) [1]. The force smoothing length is set at 1/20th of the mean inter-particle spacing. The linear power spectrum of density fluctuations was calculated by using the code CAMB (Lewis et al., 2000). Initial conditions were generated using the Zel'dovich approximation at a sufficiently high redshift $z_i = 300$ to prevent numerical artefacts (Crocce et al., 2006). The background cosmology is based on WMAP 5-year data with constraints from baryonic acoustic oscillation and high redshift supernovae, i.e. $\Omega_M = 0.27,\ \Omega_\Lambda = 0.73,\ h = 0.701,\ \sigma_8 = 0.8,\ n = 0.96$ (Komatsu et al., 2009). The simulation has the size of $47\,h^{-1} = 67$ Mpc with corresponding particle number of 1728^3 (i.e. particle mass is $1.076 \times 10^6 \mathrm{M}_\odot$) to ensure reliable halo identification down to $10^8\,\mathrm{M}_\odot$ (with 50 particles). 76 snapshots of particle data were recorded at intervals equally spaced in time, $\Delta t = 11.5$ Myr, from $z = 30$ to $z = 6$. The SPH-style smoothing scheme using nearest neighbours to transform the data to lower resolution with 306^3 or 612^3 cells in order to used in the reionisation simulation. Dixon et al. (2016) showed the convergence between the two resolutions so we will use only the coarser 306^3 cells, with the corresponding cell width of 219 kpc), in this work.

[1] See `https://github.com/jharno/cubep3m`

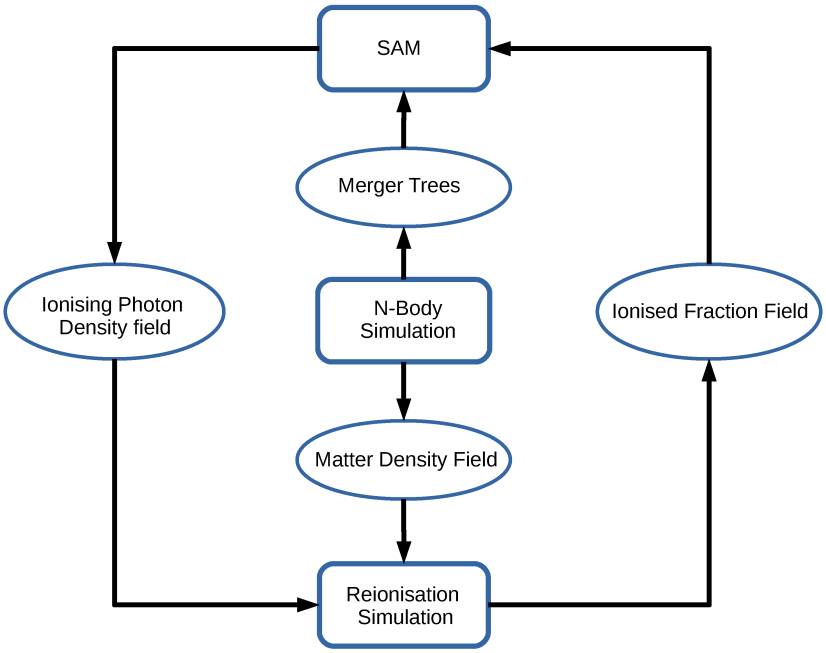

Figure 4.1: The diagram represents the combination of L-Galaxies and the semi-numerical reionisation simulation.

4.2 Halo finding and merger tree building

At each output time, the pipeline produced a friends-of-friends (FoF) catalogue by linking particles with separation less than 0.2 of the mean value (Davis et al., 1985a). We define the centre of a FoF group to be its potential minimum. The virial radius is defined as the radius of the largest sphere around this centre and a mean density exceeding 200 times the critical density of the Universe. The total mass within the virial radius is defined as the virial mass of the group. Virial radius and virial mass are then related by

$$R_{200c} = \left[\frac{G}{100}\frac{M_{200c}}{H^2(z)}\right]^{1/3}. \tag{4.1}$$

Then, the Subfind algorithm was applied to each FoF group to identify all its self gravitational bound structures, called subhaloes (Springel et al., 2001a). The most-massive self-bound subhalo in a FoF group is classified as its main subhalo (sometimes the main halo) and normally host most of its mass. Other subhaloes in the FoF group are classified to as satellite subhaloes.

The merger trees in this project are constructed by LHaloTree algorithm, which is described in the supplementary information of Springel et al. (2005b), by linking each subhalo found in a given output to only one descendent at the later time using the particles IDs which are assigned to every particles. Each particle ID is unique and fixed to only one

particle across the simulation. The further comparisons of merger tree algorithms can be found in Srisawat et al. (2013); Avila et al. (2014); Lee et al. (2014); Wang et al. (2016).

4.3 The Munich semi-analytic model of galaxy formation L-Galaxies

The Munich SAM, or L-Galaxies (Springel et al., 2001a; De Lucia et al., 2004b; Kauffmann et al., 2004; Springel et al., 2005a; Croton et al., 2006; De Lucia and Blaizot, 2007; Guo et al., 2011, 2013; Henriques et al., 2013, 2015), has been developed over decades to include most of the relevant processes that affect galaxy evolution. A brief summary of the galaxy formation recipes is in Appendix A, but we strongly encourage the readers to the supplementary material of Henriques et al. (2015), Hen15 hereafter, for more detailed descriptions. L-Galaxies has succeeded in predicting various physical properties of galaxies in the local Universe and up to redshift $z = 3$ (Henriques et al., 2013, Hen13 hereafter). It also produces results to agree reasonably well with the observations at high redshift (Clay et al., 2015). In this work, we deploy the physical recipes used in Hen15, except the baryonic infall model which is discussed in Section 4.3.1, to use with the high resolution N-body simulation described in Section 4.1.

4.3.1 Baryonic infall

We assume that collapsed dark matter haloes (FoF groups in this work) always have a mass of associated baryons given by the universal baryon fraction, $f_b = \Omega_b/\Omega_m$ (White and Frenk, 1991). This value specifies the maximum baryons a halo can possess. At any time, the matter which has not previously accreted into any objects will be added into the central objects in the form of primordial gas. However, some halos might not be able to accrete the total amount of baryons due to many factors, especially the ionising UV radiation which can heat the IGM (Efstathiou, 1992) and increase the local Jeans mass (Dijkstra et al., 2004b). We parametrise the reduction of infall gas in terms of a baryon fraction modifier, $f_{\rm mod}$, which specifies the total baryon mass a halo can capture in its lifetime. We model the infall mass,

$$m_{\rm infall} = f_{\rm mod} f_b M_{\rm 200c} - m_{\rm baryon}, \tag{4.2}$$

where

$$m_{\rm baryon} = \sum_{i=0}^{N_{\rm gal}-1} (m^i_* + m^i_{\rm cold} + m^i_{\rm hot} + m^i_{\rm ejected} + m^i_{\rm BH}), \tag{4.3}$$

and $0 \leq f_{\rm mod} \leq 1$. $N_{\rm gal}$ is the total number of galaxies inside the virial radius of the halo. m_*, $m_{\rm cold}$, $m_{\rm hot}$, $m_{\rm ejected}$, and $m_{\rm BH}$ are the total mass of stars, cold gas disc, hot atmosphere, ejected phase gas, and super massive black hole respectively.

Hen15 assumes that $m_{\rm infall}$ is non-negative. If the right hand side of Equation 4.2 is less than 0, $m_{\rm infall}$ will be set to be 0 (we hereafter call this scheme as stripping scheme "0"). $f_{\rm mod}$ is a function of $M_{\rm 200c}$, redshift and other local environment properties. In the standard version of L-GALAXIES, it only depends on $M_{\rm 200c}$ and redshift assuming homogeneous reionisation (Gnedin, 2000; Okamoto et al., 2008). The ability to use local environment properties, such as local ionised fraction, as a factor to calculate $f_{\rm mod}$ is the main feature of this work. The conditions used for $f_{\rm mod}$ will be described in detail in Section 4.4.

The excess baryon problem

According to Equation 4.2 and the assumption used in Section 4.3.1, $f_{\rm baryon} = m_{\rm baryon}/M_{\rm 200c}$ can be greater than $f_{\rm b}$ if the virial mass of a FoF group decreases over a time step, which is unlikely to happen in ΛCDM cosmology. However, many factors such as discrete snapshots, halo finding methods, and merger tree building algorithms are able to trigger this phenomenon. The data dark matter particles are stored at discrete time-steps. In order to identify the bound structures, i.e. haloes, halo finder methods are applied to the stored particle data. However, most of the available halo finder algorithms, including SUBFIND which is used in this work, use the information from only one snapshot to identify bound structures. Therefore, the temporal fluctuation of halo masses can arise any time. When combine with merger tree builders, which track haloes in simulations across time-steps, it is not uncommon to find that haloes decrease in mass at some point in their history. (Srisawat et al., 2013; Avila et al., 2014). This issue will be discussed in detail in Part III of this thesis.

By restricting $m_{\rm infall}$ in Equation 4.2 to be non-negative, the baryon contents inside haloes can be overestimated because the total baryon mass will be set to the maximum $f_{\rm mod} f_{\rm b} M_{\rm 200c}$ in their lifetime. The spike in mass history, e.g. from host-subhalo misidentification, will lead to excess baryon content. Particularly, the halos with less than a few hundred dark matter particles which the mass is less stable and prone to numerical fluctuation. Figure 4.5 shows the average baryon fraction as a function of the virial mass of FoF groups at redshift $z = 6$, 8 and 10. The coloured regions show the 68% confident intervals ($\pm 1\sigma$) of the distributions. The left panels are the plots for the stripping "0"

scheme. They show that the average baryon fraction of a haloes, with virial mass less than 10^9 $M_\odot$, using the"No suppression" model (see Section 4.5) are greater than the global baryon fraction of the Universe using this stripping scheme.

We try to fix this issue by introducing a buffer gas phase called the "excess" state and allowing $m_{\rm infall}$ to be negative. Under these circumstances, material is moved from the halo into this excess phase which therefore acts like a buffer between the primordial gas from the IGM and the baryon reservoir inside the haloes.

If $m_{\rm infall} > 0$ and $m_{\rm infall} > m_{\rm excess}$, where $m_{\rm excess}$ is the total mass of the excess state, the total mass of $m_{\rm infall} - m_{\rm excess}$ is supplied from primordial gas from the surrounding IGM and transferred into the hot gas phase. Then all the gas, including metals, in the excess state will be distributed into the hot gas and ejected phases. The total mass of $m_{\rm excess}\left(\frac{m_{\rm hot}}{m_{\rm hot}+m_{\rm ejected}}\right)$ is given to the hot atmosphere and $m_{\rm excess}\left(\frac{m_{\rm ejected}}{m_{\rm hot}+m_{\rm ejected}}\right)$ is given to the ejected state. The metal mass is also distributed at the same proportion as the mass transferred to the respecting gas states. If $m_{\rm infall} < m_{\rm excess}$, no primordial gas will be accreted from the surrounding IGM. Then, the total mass of $m_{\rm infall}\left(\frac{m_{\rm hot}}{m_{\rm hot}+m_{\rm ejected}}\right)$ and $m_{\rm infall}\left(\frac{m_{\rm ejected}}{m_{\rm hot}+m_{\rm ejected}}\right)$ from the excess mass reservoir are given to the hot atmosphere and the ejected state respectively.

If $m_{\rm infall} < 0$, the total baryon of $|m_{\rm infall}|$ will be "stripped" from the hot gas and ejected phase to the excess state. The total mass of $|m_{\rm infall}|\left(\frac{m_{\rm hot}}{m_{\rm hot}+m_{\rm ejected}}\right) \leq m_{\rm hot}$ from the hot atmosphere and $|m_{\rm infall}|\left(\frac{m_{\rm ejected}}{m_{\rm hot}+m_{\rm ejected}}\right)$ from the ejected state will be transferred to the excess state. In the case that $|m_{\rm infall}|\left(\frac{m_{\rm ejected}}{m_{\rm hot}+m_{\rm ejected}}\right) > m_{\rm ejected}$, the remaining mass, $(|m_{\rm infall}|\left(\frac{m_{\rm ejected}}{m_{\rm hot}+m_{\rm ejected}}\right) - m_{\rm ejected}) \leq m_{\rm hot}$, from the hot atmosphere will be given to the excess state. We hereafter call this the stripping "1" scheme. This scheme is expected to reduce the effect from the mass fluctuation in merger trees, especially the overestimated baryon content in low mass haloes.

4.3.2 UV Luminosity modelling

The initial mass functions (IMF) and stellar population synthesis (SPS) models are crucial parts of galaxy formation theory as they link the masses, ages and metallicities predicted for stars to the observable emission at various wavelengths. The UV luminosity of the mock galaxies in this work is generated by following Hen15 to use Chabrier IMF (Chabrier, 2003) and the publicly released but still unpublished Charlot & Bruzual (2007) as the SPS model. Since actively star-forming galaxies are known to be rich in dust which is able to significantly absorb UV light, we model the dust absorption according to Section S1.14 of

the supplementary material of Hen15.

4.4 Reionisation

We adopt the semi-numerical methods given in Majumdar et al. (2014), which is based on the excursion-set formalism described in Furlanetto et al. (2004), and similar to Zahn et al. (2007), Mesinger and Furlanetto (2007), Choudhury et al. (2009), and Santos et al. (2010).

In this work, we assume that the number of reionising photons is proportional to the number of hydrogen atoms in the forming stars inside galaxies, hence the rate of ionising photons emitted into the IGM, $\dot{N}_\gamma$, can be expressed as:

$$\dot{N}_\gamma = \Sigma_{\rm esc} \frac{\dot{m}_*}{m_{\rm HI}}, \tag{4.4}$$

where $\Sigma_{\rm esc}$ is the the number of photons per a hydrogen atom of forming stars that escape into the IGM, $\dot{m}_*$ is the gross star formation rate, and $m_{\rm HI}$ is the mass of a hydrogen atom. $\Sigma_{\rm esc}$ can be expressed as $f_{\rm esc} N_i$, where N_i is the mean number of ionising photons produced by an atom cycled through stars, averaged over the initial mass function (IMF) of the stars, and $f_{\rm esc}$ is the fraction of these ionising photons that escapes into the IGM (Haiman and Holder, 2003). In this work, we assume $\Sigma_{\rm esc}$ is a constant. In the future, we will investigate $\Sigma_{\rm esc}$ as a function of redshift, metallicity, and the stellar age of the galaxies.

L-Galaxies gives the information about locations and $\dot{m}_*$ of galaxies. We define $N_{\gamma,\rm total}$ is the total number of ionising photons that have been emitted by a galaxy and its progenitors into the IGM up to a given time, t_f:

$$N_{\gamma,\rm total} = \Sigma_{\rm esc} \int_0^{t_f} \frac{\dot{m}_*(t)}{m_{\rm HI}} dt, \tag{4.5}$$

where $\dot{m}_*(t)$ is the total star formation rate of the galaxy or its progenitors at time t.

We then follows the reionisation recipe described in Section 2.3.1 of Majumdar et al. (2014) to calculate the growth of H II bubbles. By assuming that dark matter and baryon are fully coupled with no bias, the number density field of neutral hydrogen can be estimated as

$$n_{\rm HI}(\mathbf{x}) = \frac{f_{\rm b} X_{\rm HI}}{m_{\rm HI}} \rho_{\rm DM}(\mathbf{x}), \tag{4.6}$$

where $f_{\rm b} = \Omega_b/\Omega_m$ is the universal baryon fraction, $\rho_{\rm DM}(\mathbf{x})$ is the dark matter density field, and $X_{\rm HI}$ is the mass fraction of H I in the Universe.

We then estimate the average number density of photons within a spherical radius R with the centre located at a coordinate $\mathbf{x}$, $\langle n_\gamma(\mathbf{x})\rangle_R$, and compare it to the corresponding average number density of neutral hydrogen $\langle n_{\rm HII}(\mathbf{x})\rangle_R$. The smoothing radius, R, is ranged from the grid cell size, $R_{\rm cell}$, to a radius $R_{\rm max}$, which is determined by the assumed mean free path of the photon at the concerned redshift. With $N_{\rm rec}$ as the averaged number of recombinations per hydrogen atom in the IGM, which we assume to be uniformly constant[2,3].The cell located at a coordinate $\mathbf{x}$ is considered to be ionised if there exists a radius $R \in [R_{\rm cell}, R_{\rm max}]$ such that

$$\langle n_\gamma(\mathbf{x})\rangle_R \geq \frac{\langle n_{\rm HI}(\mathbf{x})\rangle_R}{(1+N_{\rm rec})}. \tag{4.7}$$

If the condition is not satisfied, the ionised fraction of the cell is given by the maximum, over all $R \in [R_{\rm cell}, R_{\rm max}]$, of

$$x_{\rm HII}(\mathbf{x}) = \frac{\langle n_\gamma(\mathbf{x})\rangle_R}{\langle n_{\rm HI}(\mathbf{x})\rangle_R(1+N_{\rm rec})}. \tag{4.8}$$

Since we assume $\Sigma_{\rm esc}$ and the averaged number of recombinations per hydrogen atom to be constant, the factor $(1+N_{\rm rec})$ can be included within the definition of $\Sigma_{\rm esc}$. In other words, the effect of recombination can be compensated by increasing the efficiency of the sources. Thus, we define the effective $\Sigma_{\rm esc}$:

$$\Sigma_{\rm esc,eff} = \frac{\Sigma_{\rm esc}}{1+N_{\rm rec}} \tag{4.9}$$

To be able to run the simulation repetitively to finely tune the free parameters in the model, the method needs to be run in a fully parallel manner. We parallelise the calculation by assigning N different values of the smoothing radius, $R_1,\ R_2,\ R_3,\ ...,\ R_N$, distributed to $M \leq N$ processors. When the ionisation fields for a given value smoothing radius R_i, $x_{{\rm HII},R_i}(\mathbf{x})$, are processed for all $i \in \{1,2,3,...,N\}$, then the final result for the cell at a given coordinate $\mathbf{x}$;

$$x_{\rm HII}(\mathbf{x}) = \max_{1\leq i\leq N} x_{{\rm HII},R_i}(\mathbf{x}). \tag{4.10}$$

This method is proven to scale well to M processor given that the memory is sufficient.

[2]It is also possible to incorporate a self-shielding criterion in this simulation based on a density dependent recombination scheme (Equation 15 of Choudhury et al., 2009), which we do not consider in the comparisons in this work.

[3]The effect of reionisation on the recombination , e.g. the reduction of the clumping factor (Pawlik et al., 2009), is not considered in this comparison.

4.5 Suppression models

In order to investigate the effect of the suppression caused by the reionisation, we compare a simple patchy suppression model which the ionisation state of the local IGM is taken into account, against the homogeneous suppression model which is the standard model for L-Galaxies. We also include the model which does not include reionisation suppression as a reference. The detailed descriptions of all models can be found below.

4.5.1 "No suppression" model

In this model, we assume that there is no suppression from reionisation. Therefore, we set $f_{\rm mod}$ in eq. 4.2 to 1. This implies the baryon fraction inside the virial radius of any halo is the same as the universal baryon fraction.

4.5.2 "Homogeneous" model

This model assume homogeneous reionisation. Hen15 model utilises the results from Gnedin (2000) who defines a filtering halo mass, $M_{\rm F}(z)$, with a model of $f_{\rm mod}$ as a function of redshift and $M_{\rm 200c}$ of the halo;

$$f_{\rm mod}(z, M_{\rm 200c}) = \left(1 + (2^{\alpha/3} - 1)\left[\frac{M_{\rm 200c}}{M_{\rm F}(z)}\right]^{-\alpha}\right)^{-\frac{3}{\alpha}}. \tag{4.11}$$

For haloes with $M_{\rm 200c} > M_{\rm F}(z)$, suppression of the baryon fraction is small, but for haloes with $M_{\rm 200c} \ll M_{\rm F}(z)$ the baryon fraction drops to $(M_{\rm 200c}/M_{\rm F}(z))^3$. Hen15 adopts $\alpha = 2$ and take $M_{\rm F}(z)$ from the numerical results of Okamoto et al. (2008). $M_{\rm F}$ varies from $\sim 6.5 \times 10^9\, M_\odot$ at $z = 0$, to $\sim 10^7\, M_\odot$ at $z = 8$. Figure 4.2 shows the relationships between $f_{\rm mod}$ and $M_{\rm 200c}$ at $z = 6$, 7, 8, and 9.

4.5.3 "Patchy" suppression model

We model a simple patchy suppression in the same manner as the mass-dependent suppression of LMACHs (gS) model described in Section 3.2 of Dixon et al. (2016) to modify $f_{\rm mod}$ in Equation 4.2.

For an FoF group located at $\mathbf{x}$, if the grid cell containing the centre of the FoF group is considered to be ionised, $x_{\rm HII}(\mathbf{x}) > 0.5$ in this work,

$$f_{\rm mod} = \begin{cases} 0 & \text{if } M_{\rm 200c} < 10^8 {\rm M}_\odot \\ \log_{10}\left(M_{\rm 200c}/10^8\ {\rm M}_\odot\right) & \text{if } 10^8 {\rm M}_\odot \leq M_{\rm 200c} < 10^9 {\rm M}_\odot \\ 1, & \text{if } M_{\rm 200c} \geq 10^9 {\rm M}_\odot. \end{cases} \tag{4.12}$$

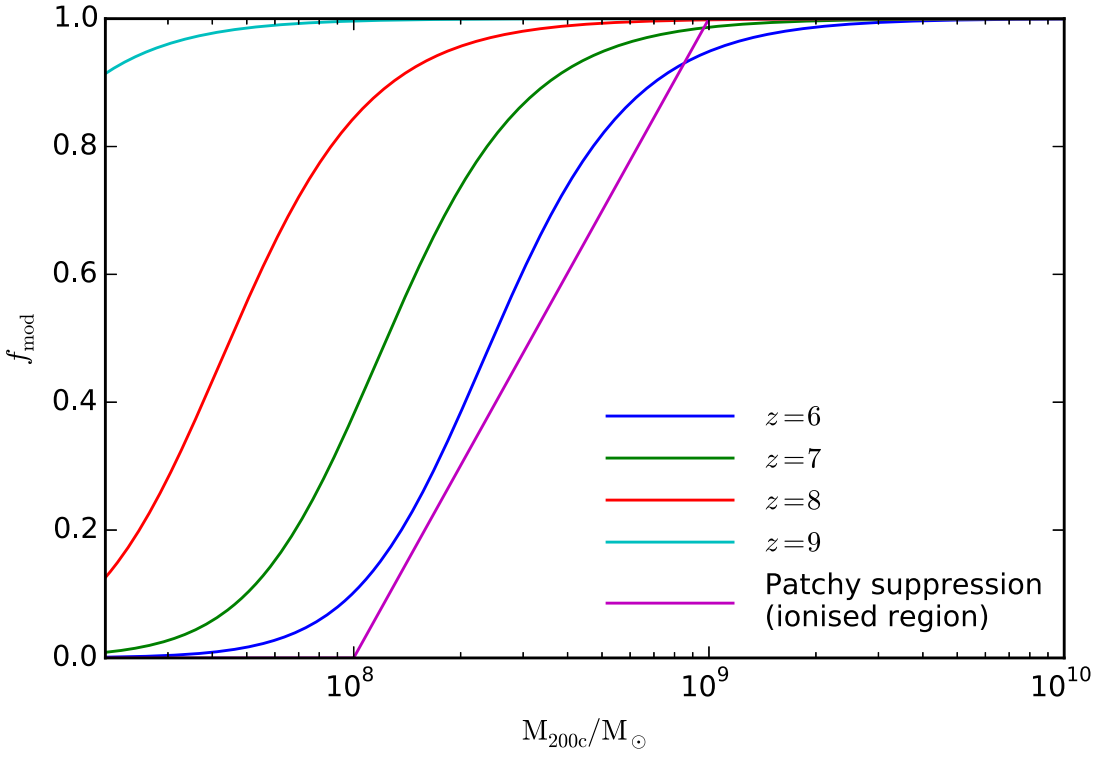

Figure 4.2: The the relationship between $f_{\rm mod}$ and M_{200c} of "patchy" model in ionised regions, compared to the relationship between $f_{\rm mod}$ and M_{200c} from Okamoto et al. (2008) at $z = 6$, 7, 8, and 9.

If the the grid cell is neutral, $x_{\rm HII}(\mathbf{x}) \leq 0.5$ in this work,

$$f_{\rm mod} = 1, \text{ for all mass range.} \tag{4.13}$$

Figure 4.2 shows the relationship between $f_{\rm mod}$ and M_{200c} of the patchy suppression model in ionised regions, compared to the model described in Okamoto et al. (2008) at different redshifts, $z = 6$, 7, 8, and 9.

4.6 Combination of L-Galaxies and the semi-numerical reionisation simulation

In order to couple galaxy evolution and reionisation in this work, we employ the algorithm illustrated in Figure 4.1 which can be described below:

1. Start from the N-body simulation, merger trees are built from the halo catalogues extracted from the simulation. All snapshots, $S_0, S_1, S_2, ..., S_{N-1}$, where N is total number of the stored snapshots, are sorted by their expansion rates. The density

field of dark matter particles of each snapshot is generated onto grids, then they can be converted to the number density fields of neutral hydrogen using Equation 4.6.

2. To produce the galaxy catalogue of a snapshot S_i, the merger trees and the ionised fraction fields of the snapshot $S_0, ..., S_{i-1}$ are supplied to L-Galaxies. For any snapshots S_i where $i < 0$, which do not exist, the simulation box is assumed to be entirely neutral.

3. To produce the integrated ionising photon field of a snapshot S_i, the galaxy catalogue at S_i is used to supply the integrate star formation and average onto grids. The relationship in Equation 4.5 is applied to convert the integrated star formation to the amount of the integrated ionising photons.

4. To produce the ionised fraction field of a snapshot S_i, the integrated ionising photon field and the number density field of neutral hydrogen of the snapshot S_i are supplied to the reionisation simulation (see Section 4.4).

5. Repeat step 2-4 for all snapshots, from S_0 to S_{N-1}.

4.7 Model calibration

We make use of the UV luminosity function at $z = 6,\ 7$, and 8 described in Bouwens et al. (2015b), which utilises the combination of data from HUDF09, HUDF12, ERS, BoRG/HIPPIES, and CANDELS programs on HST, to calibrate the free parameters in L-Galaxies model. The dataset contains a sample of more than 10000 galaxy candidates at $z \geq 4$, which resolves 857, 481, and 217 galaxies at $z \sim 6$, $z \sim 7$, and $z \sim 8$ respectively. The calibrating engine is the MCMC fitting engine in L-Galaxies, which is able to search for the coordinate in high-dimensional parameter space to make the best fit to a given set of observations (Henriques et al., 2009).

In Figure 4.3, we show the UV luminosity functions of the models at redshift $z = 6$, 7, and 8, along with the constraining observations (Bouwens et al., 2015a) and some recent observations, i.e. Bouwens et al. (2007, 2010a); Oesch et al. (2010); McLure et al. (2011); Duncan et al. (2014); Bowler et al. (2014) at redshift $z = 6,\ 7$, and 8. The error bars on the observed data points were contributed from Poisson noise and the uncertainties in photometric redshift determination provided in the publications. We are able to achieve an excellent match to the normalisation, shape, and evolution of the observed UV luminosity function across all plotted redshifts. We found that there is no significant discrepancy in

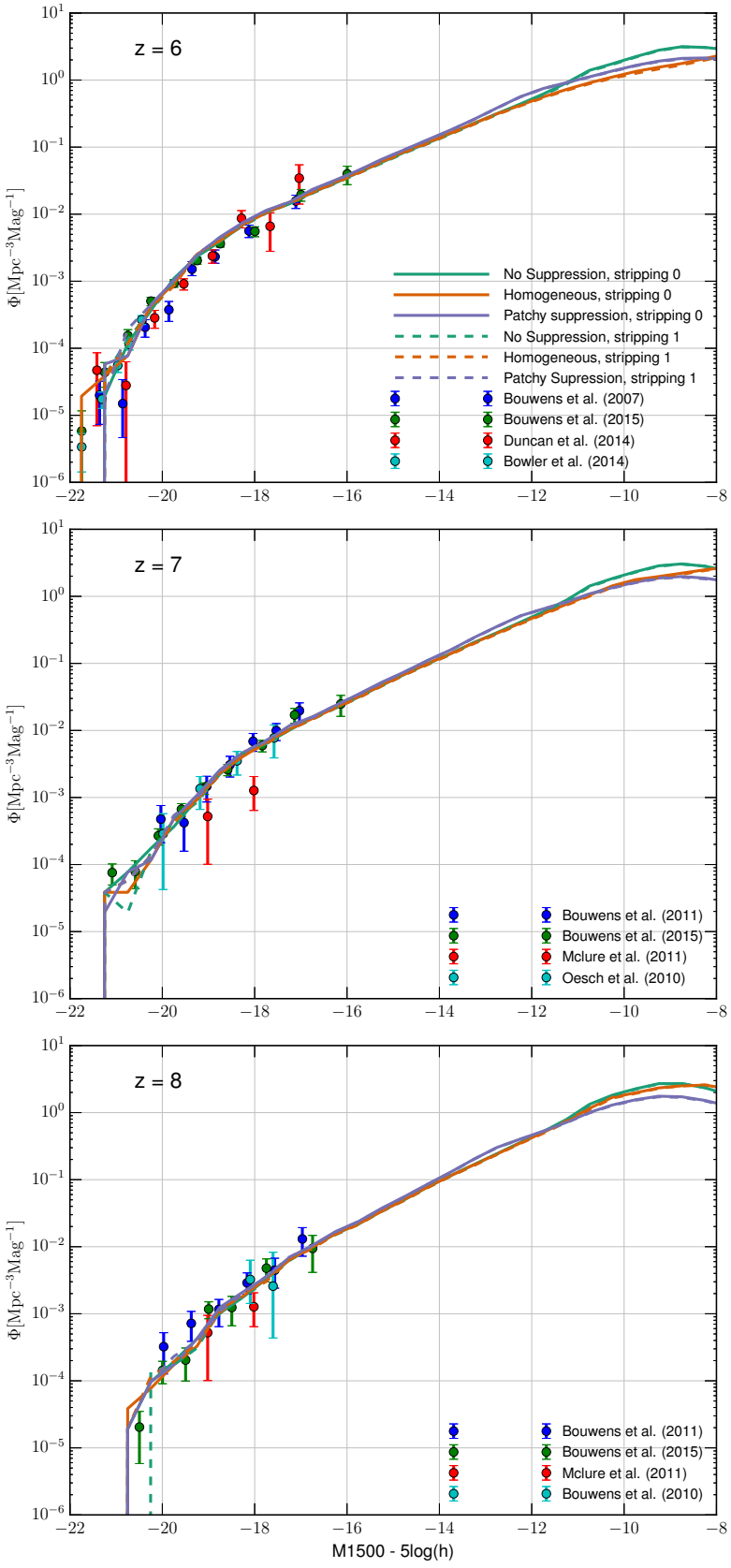

Figure 4.3: The UV luminosity function of galaxies from different models at $z = 6$ (top), $z = 7$ (middle), and $z = 8$ (bottom).

Table 4.1: The parameters, with the corresponding descriptive equations in the supplementary material of Henriques et al. (2015), used in L-Galaxies in this work compared to Henriques et al. (2013), Hen13, which assumed 7-year Wilkinson Microwave Anisotropy Probe (WMAP7) cosmology and Henriques et al. (2015), Hen15, which assumed 1-year PLANCK cosmology.

Parameter	Hen13	Hen15	this work	unit
$\alpha_{\rm SF}$(SF eff - eq. S14)	0.055	0.025	0.18	
$m_{\rm crit,0}$(Gas mass threshold - eq. S15)	0.38	0.24	0.20	$10^{10}{\rm M}_\odot$
$\alpha_{\rm SF,burst}$(SF burst eff - eq. S33)	0.56	0.60	0.50	
$\beta_{\rm SF,burst}$(SF burst slope - eq. S33)	0.70	1.9	0.56	
$k_{\rm AGN}$(Radio feedback eff - eq. S24)	-	5.3×10^{-3}	3.0×10^{-3}	${\rm M}_\odot{\rm yr}^{-1}$
$f_{\rm BH}$(BH growth eff - eq. S23)	0.015	0.041	0.015	
$v_{\rm BH}$(Quasar growth scale - eq. S23)	280	750	280	km s^{-1}
ϵ (Mass-loading eff - eq. S19)	2.1	2.6	2.1	
$v_{\rm reheat}$ (Mass-loading scale - eq. S19)	405	480	405	km s^{-1}
β_1 (Mass-loading slope - eq. S19)	0.92	0.72	0.92	
η (SN ejection eff - eq. S17)	0.65	0.62	0.15	
$v_{\rm eject}$ (SN ejection scale - eq. S17)	336	100	336	km s^{-1}
β_2 (SN ejection slope - eq. S17)	0.46	0.80	0.36	
γ (Ejecta reincorporation - eq. S22)	1.8×10^{10}	3.0×10^{10}	1.8×10^{10}	
$m_{\rm r.p.}$ (Ram-pressure threshold)	0.0	1.2×10^{4}	0.0	$10^{10}{\rm M}_\odot$
$R_{\rm merger}$ (Major-merger threshold)	0.3	0.1	0.3	
$\alpha_{\rm friction}$ (Dynamical friction - eq. S32)	2.0	2.5	2.0	
y (Metal yield)	0.047	0.046	0.047	
$\Sigma_{\rm esc,eff}$ (Effective ionising eff - Equation 4.9)	-	-	2523	

the UV luminosity functions, within the ranges of the constraining data set, between all suppression models using the same set of L-Galaxies parameters. Hence, we will use only one set of parameters to apply to all suppression models for each stripping scheme. We deliberately plot the UV luminosity function to the lower brightness end to see in what range of the absolute luminosity that the discrepancy between the model appears. We can see that there is no significant difference between the "No Suppression" and the "Homogeneous" models when the absolute UV magnitude $M_{1500} - 5\log h < -14$ and no difference between the "No Suppression" and the "Patchy" models when the absolute UV magnitude $M_{1500} - 5\log h < -12$.

Table 4.1 shows the parameters used in L-Galaxies in both stripping schemes in this work compared to the work in Hen13 and Hen15. In both stripping schemes, the calibrated parameters are similar to those used in Hen13, which assumed the cosmology from 7-year Wilkinson Microwave Anisotropy Probe (WMAP7), with larger star formation rate efficiency, $\alpha_{\rm SF}$, which is 0.18 compared to 0.055 in Hen13 and 0.025 in Hen15. We also found smaller value of supernova ejection efficiency, η, which is 0.15 compared to 0.65 in Hen13 and 0.62 in Hen15. The parameters provides an excellent match with the constraining observational data across redshifts $z = 6,\ 7$, and 8. Due to the relatively small size of the simulation, the number density of the galaxies of UV magnitude less than -21 at $z = 6$, -20 at $z = 7$, and -19 at $z = 8$ diverges from the constraining observations, but still reside within the large uncertainty regimes from the other observations.

In addition to the UV luminosity function at high redshifts, we use the measurement of integrated optical depth to Thomson scattering of CMB photons by free electrons from 3-year data of *Planck* satellite (Planck Collaboration et al., 2015), to constrain $\Sigma_{\rm esc,eff}$ in Equation 4.9, which determines the total ionising photons from galaxies. By assuming that the fraction of the ionised helium is negligible, the integrated optical depth of electron scattering, τ_e, can be calculated as:

$$\tau_e(z_0) = \int_0^{z_0} \frac{cdt}{dz}(1+z)^3 \sigma_T x_{\rm HII}^m \langle n_{\rm H}\rangle dz, \tag{4.14}$$

where $\sigma_T = 6.652 \times 10^{-25}$ cm^2 is the Thomson scattering cross section, $x_{\rm HII}^m$ is the mass-weight global ionised fraction of hydrogen, and the average co-moving density of hydrogen $\langle n_{\rm H}\rangle = 1.88 \times 10^{-7}(h^2\Omega_b/0.022)$ cm^{-3}.

For comparison, we use the same $\Sigma_{\rm esc,eff}$ for every suppression model in each stripping scheme. Figure 4.4 shows the electron scattering optical depths of the models used in this work, against the observational measurement from the *Planck* satellite (Planck Collaboration et al., 2015). By using the calibrated sets of parameters in Table 4.1, we do not

notice more than 1% difference in the integrated optical depth of electron scattering, τ_e, across different suppression models in both stripping schemes.

Despite of numerous impressive efforts in both observational and theoretical studies to quantify $f_{\rm esc}$, it remains largely uncertain, from one percent to unity. While there is no consensus on the value of $f_{\rm esc}$, there are estimations from various methods. For example, Bolton and Haehnelt (2007) used a large set of hydrodynamic simulations to estimate $f_{\rm esc} \sim 20 - 30\%$ at $z \sim 6$; Finkelstein et al. (2012a) estimated average $f_{\rm esc}$ to be $\sim 30\%$ in order to get a fully ionized IGM at $z = 6$; Kuhlen et al. (2012a) found a strong redshift evolution of escape fraction, increasing from $\sim 4\%$ at $z = 4$ to unity at higher redshifts, in order to simultaneously satisfy reionisation and lower redshift Lyman-α forest constraints; Mitra et al. (2015) used the high-redshift galaxy luminosity function data to estimate $f_{\rm esc} \sim 10\%$ at $z = 6-9$; Khaire et al. (2016) studied the contribution of QSOs and galaxies to the hydrogen ionising emissivity to constrained $f_{\rm esc} \sim 0.14 - 0.22$ at $z > 5.5$. Our best estimated $\Sigma_{\rm esc,eff}$ to reproduce the observed value of the electron scattering optical depths for both stripping "0" and "1" schemes is 2523. Depending on the IMF and the average number of hydrogen recombination, $N_{\rm rec}$, the corresponding escape fraction can be anything from a few percents (a top-heavy IMF with very low recombination) to tens of percent. If we assume Chabrier IMF with the star mass ranged between $0.001 - 120{\rm M}_\odot$, which provides $N_i \sim 50000$ (Topping and Shull, 2015), the corresponding escape fraction will be ~ 0.10 for $N_{\rm rec} = 1$ and ~ 0.15 for $N_{\rm rec} = 2$.

4.8 Results and discussions

4.8.1 Baryon fraction in haloes

The baryon content used in this work is defined in Equation 4.3. The baryon fraction in a Type 0 galaxy can be expressed as:

$$f_{\rm baryon} = \frac{m_{\rm baryon}}{M_{\rm 200c}}, \tag{4.15}$$

where $m_{\rm baryon}$ (see Equation 4.3) and $M_{\rm 200c}$ are the total baryon contents inside the virial radius and the virial mass of the host halo (FoF group in this work) the galaxy resides in. Figure 4.5 shows the averaged baryon fractions as a function of the virial mass of FoF groups at redshift 6 (top), 8 (middle), and 10 (bottom). The coloured shades show the 68% confident intervals of the distributions. The black dashed lines are the models described in Okamoto et al. (2008) at the respective redshifts. The corresponding mass-weighted average H II fraction at $z = 6$, 8, and 10 are 100%, 70%, and 17% respectively.

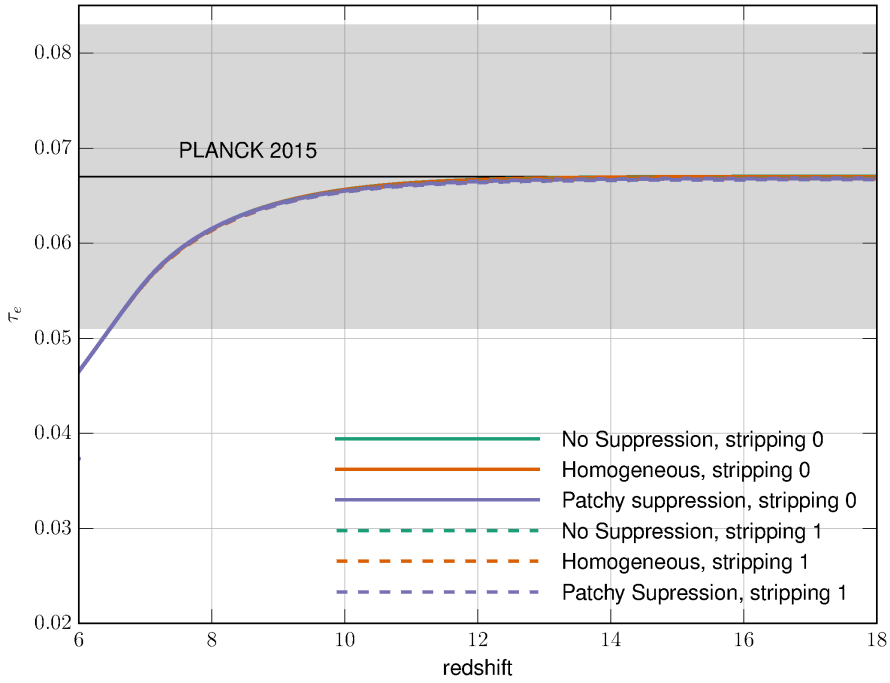

Figure 4.4: The integrated free electron scattering optical depth, τ_e, as a function of redshift. The grey horizontal line and shaded region indicate the constraints on τ_e to $z \sim 1100$ from the *Planck* 2015 data release (Planck Collaboration et al., 2015).

The figure shows that the stripping "0" scheme overestimates the baryon contents inside haloes. the "No Suppression" model gives the average baryon fractions of the haloes with virial mass less than $\sim 10^9 \mathrm{M}_\odot$ about 5% less than the universal baryon fraction at $z = 6$ and 8. Those with virial mass less than $\sim 10^{8.5} \mathrm{M}_\odot$ at $z = 10$ also display the baryon fraction which is 8% larger than the universal baryon fraction. the "Homogeneous" model also produces an average baryon fraction which is larger than the model from Okamoto et al. (2008) at all plotted redhshifts, especially for the haloes of $M_{200\mathrm{c}} = 10^8 \mathrm{M}_\odot$ at $z = 6$ which contain more than twice the expected value from the model. Note that we use Okamoto et al. (2008) model to calculate the amount of infall for all galaxies. This suggests that the stripping scheme is sensitive to the fluctuation in halo histories and prone to overestimating the baryon contents in galaxies. In this stripping scheme, the "Patchy" model shows more suppression power on those haloes with virial mass less than $\sim 10^9 \mathrm{M}_\odot$ than that of the "Homogeneous" model at all plotted redshifts. Compared to the model from Okamoto et al. (2008), this model estimates $\sim 60\%$ lower baryon fraction for $M_{200\mathrm{c}} = 10^8 \mathrm{M}_\odot$ haloes at $z = 10$ and $\sim 50\%$ at $z = 8$. However, the "Patchy" model agrees quite well with Okamoto et al. (2008) model at $z = 6$.

The stripping "1" scheme is able to constrain the average baryon fraction not to exceed

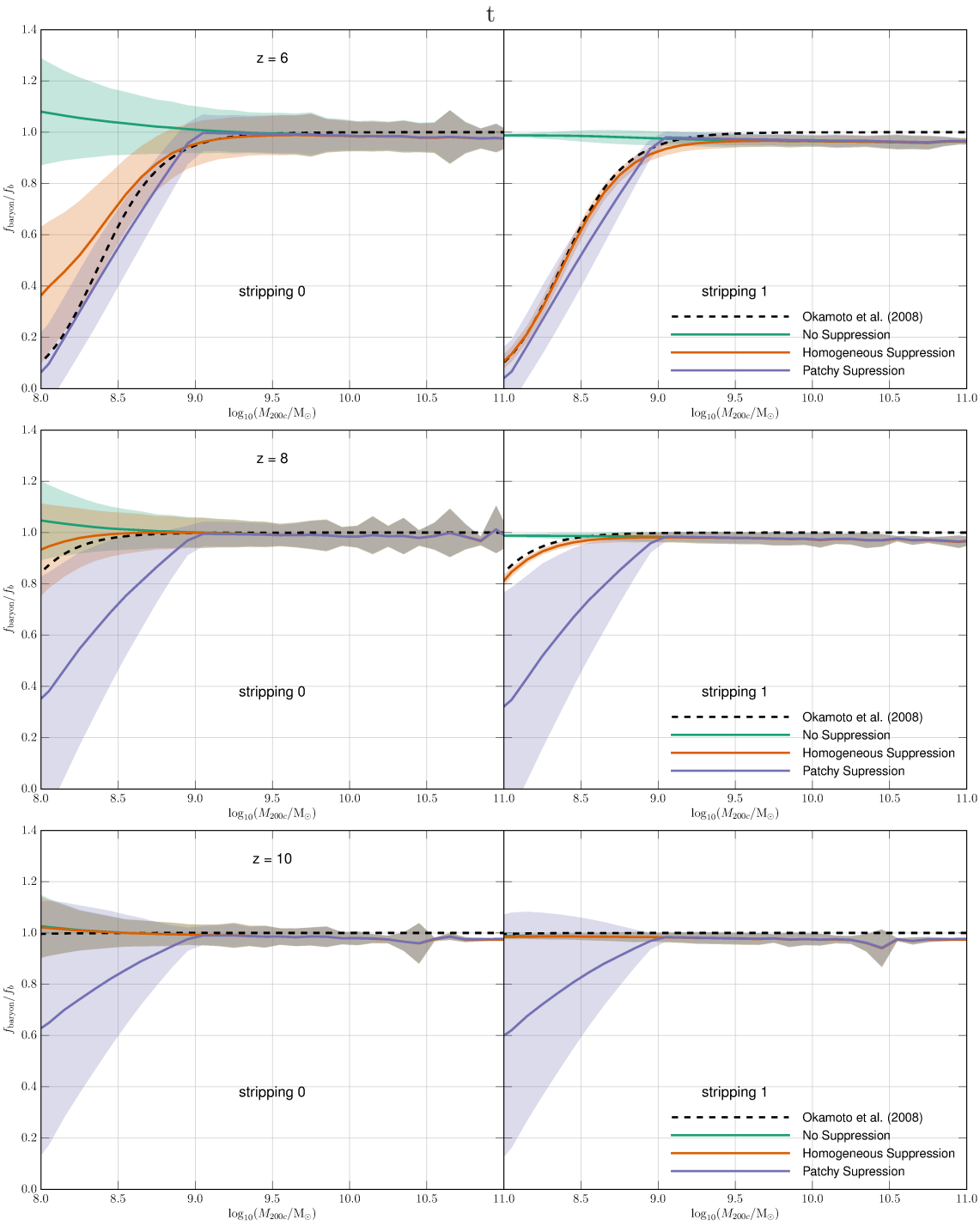

Figure 4.5: The averaged baryon fractions as a function of the virial mass of FoF groups at redshift 6 (top), 8 (middle), and 10 (bottom). The coloured shades show the 68% confident intervals of the distributions. The black dashed lines are the models described in Okamoto et al. (2008) at the respective redshifts.

the universal baryon fraction produced for every models at all redshifts. By using this stripping scheme, The average baryon fraction from the "Homogeneous" model matches Okamoto et al. (2008) model. The difference in average baryon fraction from the "Patchy" model between two stripping scheme is less noticeable, only a few percents apart.

We take some notes on the size of the uncertainty in the average baryon fractions of both stripping schemes. The standard deviation of the average baryon fraction produced by the stripping "1" scheme is much smaller than those produced by the stripping "0" scheme. For example, the uncertainties for $M_{200c} = 10^8 M_\odot$ at $z = 6$ produced by the "Homogeneous" model are 30% and 0.05% in stripping "0" and "1" schemes respectively. However, the uncertainties of those produced by the "Patchy" model remain almost identical across the stripping schemes.

We further investigated the contribution of baryonic components: hot gas, cold gas, stars, and ejected mass, to the total baryon fraction of the galaxies (see the supplementary material of Hen15 for details). Figure 4.6 shows the average mass fraction of the hot gas (first row), cold gas (second row), stellar mass (third row), and ejected mass (forth row), at $z = 6$ (left) and $z = 8$ (right). The key features of the plot can be summarised below.

- For the haloes with virial mass less than $\sim 10^{8.2} M_\odot$, which is corresponding to the virial temperature $T_{vir} \sim 10^4$ K, the hot gas cannot be cooled and form stars efficiently. Hence, most of the baryon content remains in hot gas state. The plot of hot gas fraction at $z = 6$ (first row, left panel) shows that the suppression model greatly affects the amount of hot gas in these small haloes, similar to what we have seen in the total baryon fraction in Figure 4.5.

- For haloes with virial temperature $T_{vir} > 10^4$ K, the baryon can be cooled and form stars efficiently. However, the feedback from supernovae expels the baryons in the hot atmosphere into the ejected mass state. The ejection largely minimises the effect of reionisation feedback on the amount of the hot gas, hence the star formation, as we can see the fraction of stellar mass is hardly affected by the suppression model at both redshifts. However, the cooling model in this work assumes the equilibrium cooling function for collisional processes from Sutherland and Dopita (1993) which only depends on the metallicity and temperature of the gas but ignores radiative ionisation effects. Efstathiou (1992) expected the suppression of cooling rate for $T > 10^4$ K from the photoionisation, which might increase the effect from the reionisation feedback in dwarf galaxies.

- The "Patchy" model displays its effects on the star formation in haloes with $M_{200c} >$

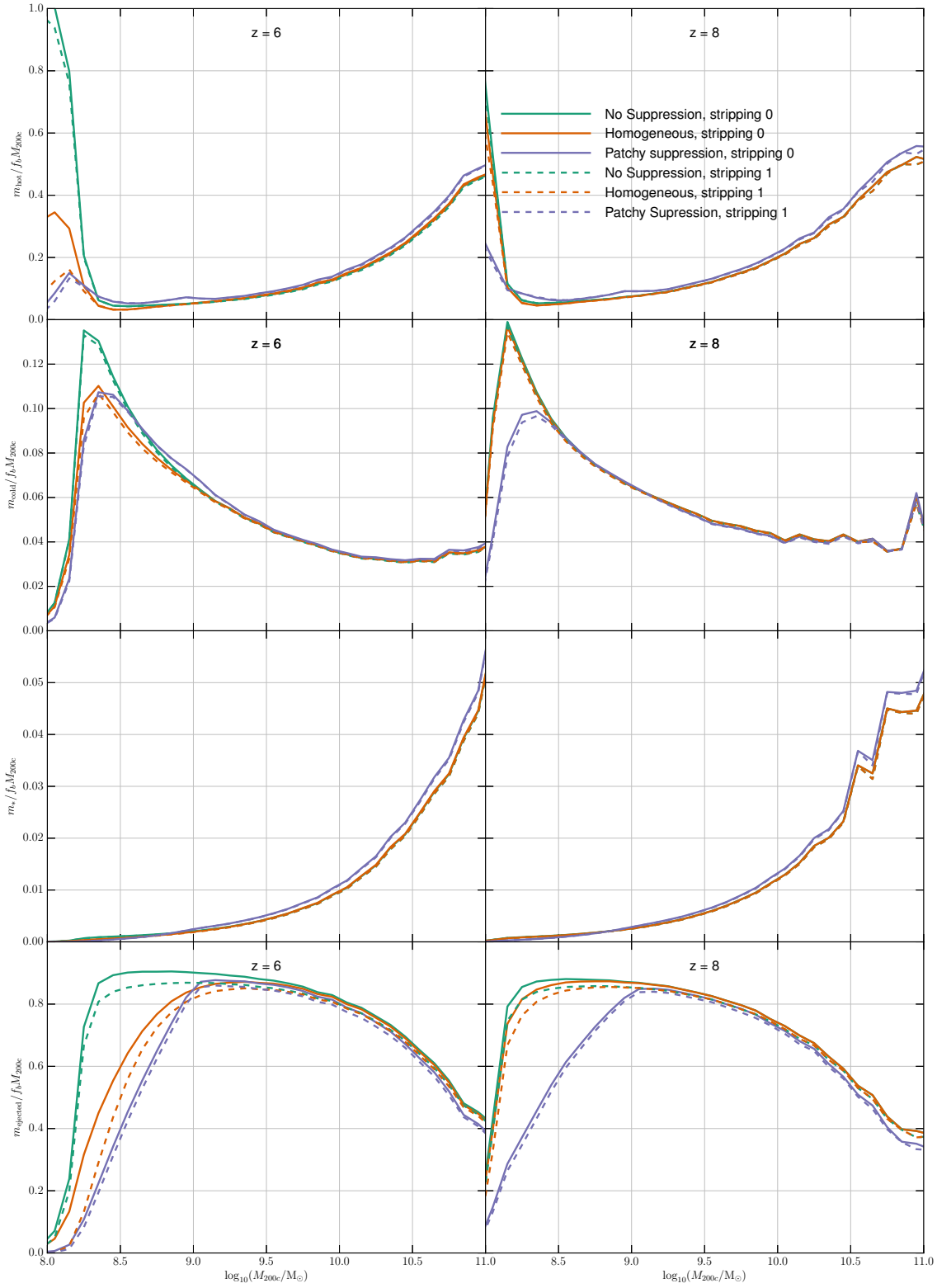

Figure 4.6: The average mass fraction of the hot gas (first row), cold gas (second row), stellar mass (third row), and ejected mass (forth row), at $z = 6$ (left) and $z = 8$ (right).

$10^9 M_\odot$, where the haloes are not suppressed in the ionised regions. The mass fractions of hot gas and star are greater than other suppression model by about 10% for $M_{200c} = 10^{11} M_\odot$. This is because the "Patchy" model assumes that the infall into the low mass haloes in ionised regions is suppressed due to the ionising background. When they are accreted into larger systems, their new hosts are able to use the baryons which have not been used in the previous time to fuel the new star formation. In the "No Suppression" model, this portion or baryons is accreted into those small haloes then cooled and condensed into galaxies. When these haloes are accreted into larger systems, their baryons remain locked up in the infalling satellite, making them unusable for the Type 0 galaxies to supply the star formation.

4.8.2 Star formation rate distribution function

Figure 4.7 shows the star formation rate distribution function (SFRDF) at $z = 6$ (left) and 7 (right) as predicted by the models along with the measurements from Smit et al. (2012) and Duncan et al. (2014) at $z = 6$, where the IMF is converted to Chabrier IMF if necessary.

Both stripping schemes produce the SFRDF's which have a remarkable agreement with Duncan et al. (2014) at redshift 6 and 7. The results from all suppression models are generally identical. At $z = 7$, the uncertainty of the data from Duncan et al. (2014) gets significantly larger than those from $z = 6$ and the data converges to those of Smit et al. (2012), which agrees very well with the SFRDF predicted by all suppression models from both stripping schemes.

4.8.3 Stellar mass

Figure 4.8 shows the stellar mass function (SMF) at $z = 6$ (left) and 8 (right) as predicted by the models along with the observational estimates from González et al. (2011), Duncan et al. (2014), and Song et al. (2016). It is important to note that the observationally derived mass functions presented in Figure 4.8 used different $m_* - M_{\rm UV}$ relationships to construct the galaxy stellar mass functions. The effect of nebular emission, which can strongly affected the estimated stellar mass (Wilkins et al., 2013), was included in Duncan et al. (2014) and Song et al. (2016) but not in González et al. (2011). The accuracy/precision of stellar mass estimates are also affected by the lower sensitivity and angular resolution of the Spitzer/IRAC imaging.

The stellar mass function from all suppression model in both stripping "0" and "1"

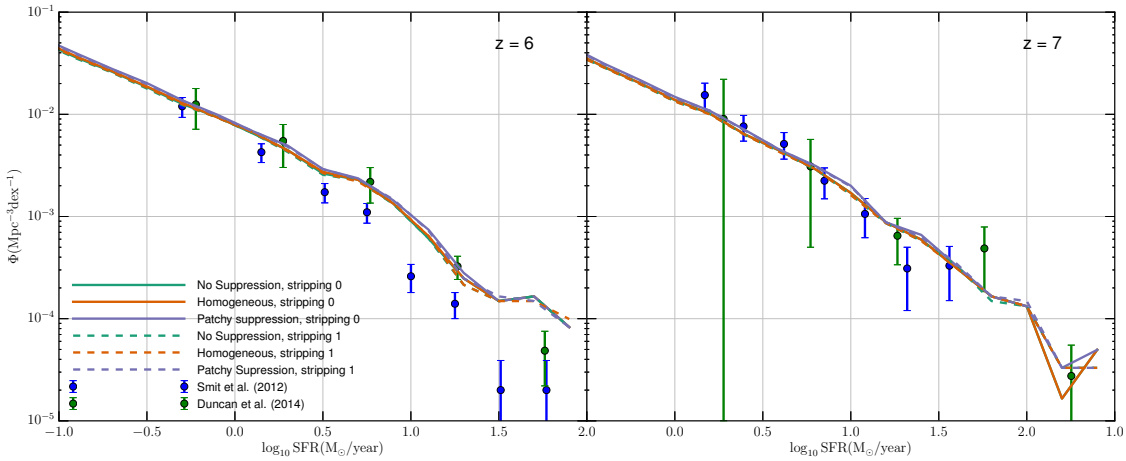

Figure 4.7: The star formation rate distribution functions (SFRDF) as predicted by the models at $z = 6$ (left) and $z = 7$ (right), along with the measurements from Smit et al. (2012) (the IMF is converted to Chabrier IMF) and Duncan et al. (2014).

schemes agree well with Song et al. (2016) at $z = 6$ and 7. Compared to Song et al. (2016)Duncan et al. (2014) employed shallower slope of $m_* - M_{\rm UV}$ relation which translates into a higher normallisation of the SMF. Similar to the SFRDF, we do not notice any significant discrepancy in the stellar mass function between the suppression models in both stripping schemes within the current observable range. The discrepancy between the suppression models is only noticeable for the galaxies with stellar mass less than $\sim 10^6 {\rm M}_\odot$. The number galaxies with stellar mass $10^{4.5} {\rm M}_\odot$ produced from the "Homogeneous" model and the "Patchy" model are $\sim 50\%$ and $\sim 70\%$ respectively less than that of the "No Suppression" model at $z = 6$, and $\sim 15\%$ and $\sim 50\%$ respectively at $z = 7$. More importantly, there is no detectable difference between the stripping schemes at all redshift using the same set of parameters. This means even the stripping scheme "0" overestimates the baryon contents inside galaxies, it is still able to deliver the stellar mass functions which are identical to those of the stripping "1" scheme.

The population of the galaxies with stellar mass between 10^6 and 10^{10} ${\rm M}_\odot$ in this work is expected to be lower than the work in Mutch et al. (2015), especially for low mass galaxies. At $z = 6$, the models in this work predict less than a half of the population density of the galaxies with $m_* \sim 10^8 {\rm M}_\odot$ estimated by Mutch et al. (2015).

Figure 4.9 shows the average total stellar mass (i.e. summed over all galaxies) in each halo (FoF group) at $z = 6$, 8, 10, 12, 14, and 16 as a function of virial mass of the hosting halo, $M_{\rm 200c}$. The coloured shades indicates the 68% confidence intervals. The dashed lines on the left panels show the fitted function $m_* \propto M_{\rm 200c}^{1.7}$. It shows excellent

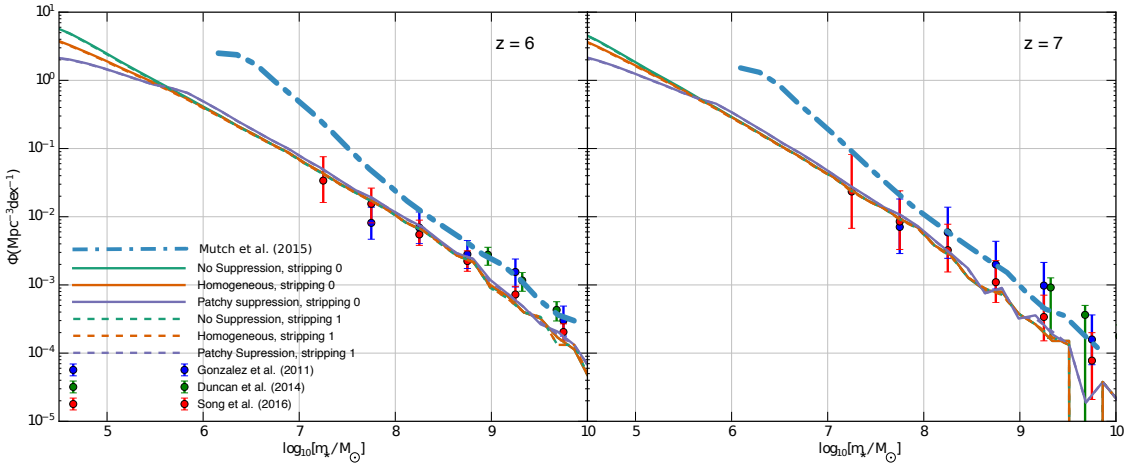

Figure 4.8: The stellar mass functions, SMF, at $z = 6$ (left) and 7 (right) as predicted by the models along with the observational estimates from González et al. (2011), Duncan et al. (2014), and Song et al. (2016), together with the SMF of both redshifts predicted by Mutch et al. (2015).

agreement with simple energy conservation arguments which suggest a slope of ~ 1.7 for supernova feedback-regulated galaxy growth and a fixed cold gas mass fraction (Wyithe and Loeb, 2013), including the result from a large sample of local galaxies to obtain the relationship $m_* \propto M_{\rm halo}^{5/3}$ for $m_* < 3 \times 10^{10}{\rm M}_\odot$, where $M_{\rm halo}$ is the mass of dark matter haloes (Kauffmann et al., 2003). We also found the redshift dependent evolution of the normalisation (see Wyithe and Loeb, 2003), which can be expressed as an empirical relation:

$$\log_{10}(m_*/{\rm M}_\odot) = 1.7\log_{10}(M_{\rm 200c}/{\rm M}_\odot) + az + b, \tag{4.16}$$

where the best estimated values for a and b from our result are -10 and $1/24$ respectively. At $z = 6$, the uncertainty of the total stellar mass of the halo with $M_{\rm 200c} < 10^9{\rm M}_\odot$ increases due to a combination of supernova and reionization feedback effects. The increase in the total stellar mass of the haloes with $M_{\rm 200c} \sim 10^8{\rm M}_\odot$ due to the lack of supernova feedback as we have seen in Figure 4.6.

4.8.4 Integrated star formation

We define the integrated star formation of a galaxy over the cosmic time, $m_{*,\rm gross}(t_f)$, as the total star formed in the galaxy and its progenitors up to a specific time t_f;

$$m_{*,\rm gross}(t_f) = \int_0^{t_f} \dot{m}_*(t)dt, \tag{4.17}$$

where $\dot{m}_*(t)$ is the total star formation rate of the galaxy or its progenitors at time t.

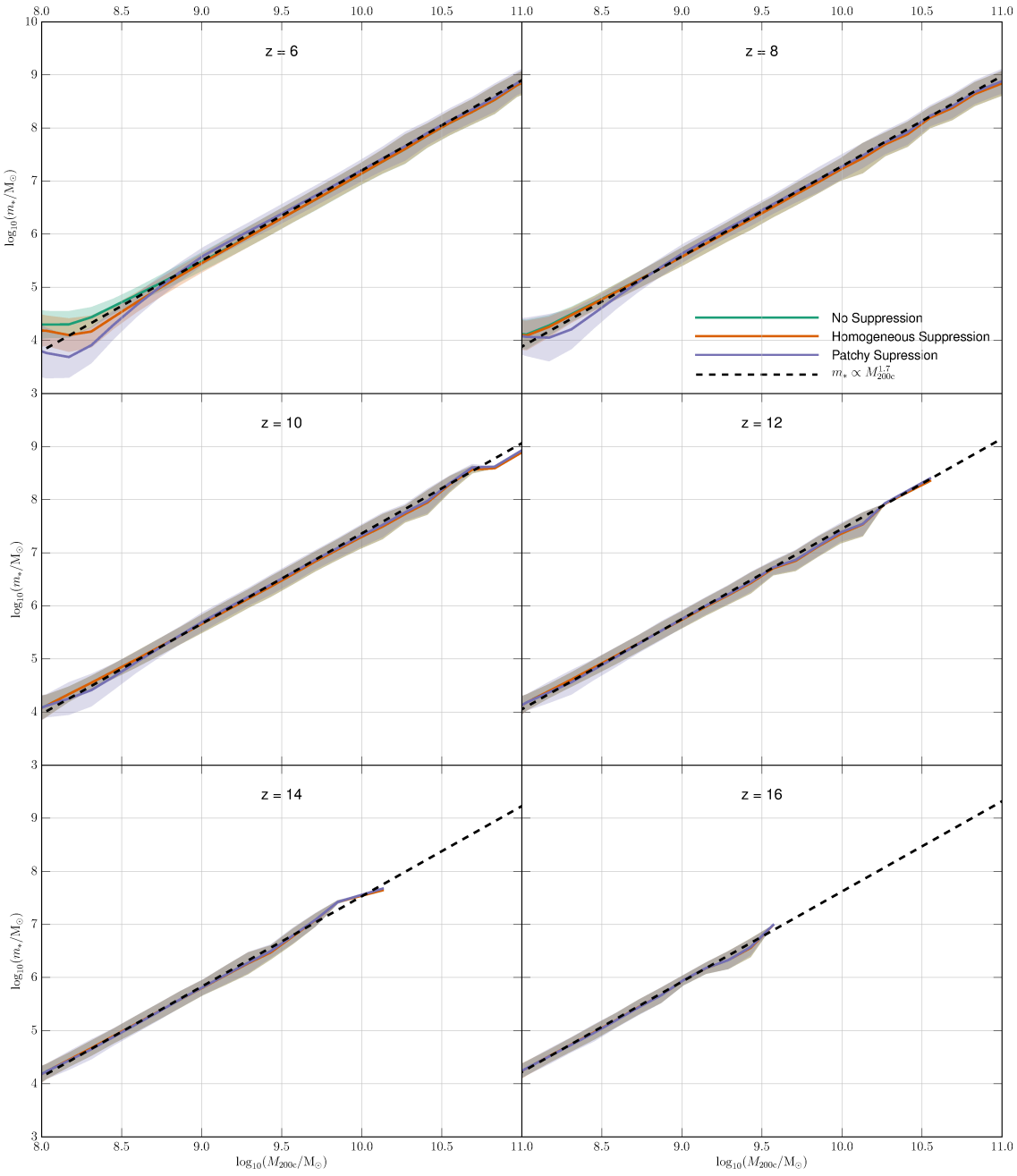

Figure 4.9: The average stellar mass, m_*, in each FoF group at $z = 6$, 8, 10, 12, 14, and 16, as a function of the virial mass of the hosting FoF group, M_{200c}. The coloured shades indicates the 68% confidence intervals. The dashed lines on the left panels show the fitted function $m_* \propto M_{200c}^{1.7}$. Both stripping schemes show identical relationship between m_* and M_{200c} for each suppression model so we will show only the plots for one scheme.

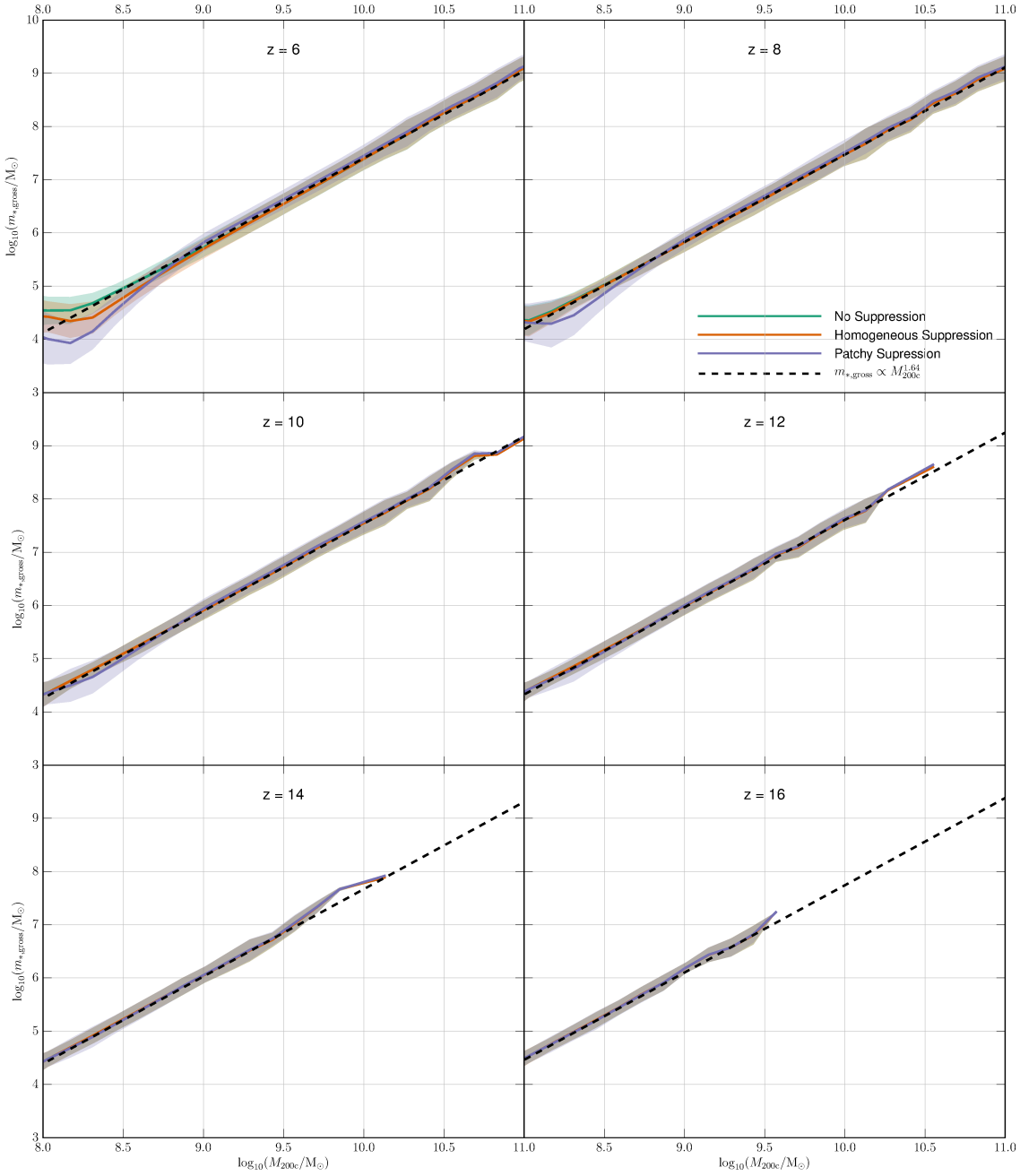

Figure 4.10: The average integrated star formation, $m_{*,\mathrm{gross}}$ up to z = 6, 8, 10, 12, 14, and 16, as a function of the virial mass of the hosting FoF group, M_{200c}. The coloured shades indicates the 68% confidence intervals. The dashed lines on the left panels show the fitted function $m_{*,\mathrm{gross}} \propto M_{200c}^{1.64}$. Both stripping schemes show identical relationship between $m_{*,\mathrm{gross}}$ and M_{200c} for each suppression model so we will show only the plots for one scheme.

Figure 4.10 shows the average integrated star formation, $m_{*,\mathrm{gross}}$ up to $z = 6$, 8, 10, 12, 14, and 16, as a function of the virial mass of the host FoF group, $M_{200\mathrm{c}}$. The coloured shades indicates the 68% confidence intervals. The dashed lines on the left panels show the fitted function $m_{*,\mathrm{gross}} \propto M_{200\mathrm{c}}^{1.64}$. Both stripping schemes show identical relationship between $m_{*,\mathrm{gross}}$ and $M_{200\mathrm{c}}$ for each suppression model so we will show only the plots of one scheme. The plots show a very linear relationship between $\log_{10}(m_{*,\mathrm{gross}}/\mathrm{M}_\odot)$ and $\log_{10}(M_{200\mathrm{c}}/\mathrm{M}_\odot)$ unless the galaxies are under the suppression from reionisation and supernovae. The top left plot shows the average $\log_{10}(m_{*,\mathrm{gross}}/\mathrm{M}_\odot)$ at a function of $\log_{10}(M_{200\mathrm{c}}/\mathrm{M}_\odot)$ at redshift 6. There is a trace of the suppression mechanism on the haloes with $M_{200\mathrm{c}} < 10^9\mathrm{M}_\odot$. The average integrated star formation of the"Homogeneous" and the "Patchy" models are 50% and 75% lower than the "No Suppression" model respectively for $M_{200\mathrm{c}} \sim 10^{8.2}\mathrm{M}_\odot$.

The linear relationship between $\log_{10}(m_{*,\mathrm{gross}}/\mathrm{M}_\odot)$ and $\log_{10}(M_{200\mathrm{c}}/\mathrm{M}_\odot)$ without the suppression from reionisation can be expressed the same way as Equation 4.16:

$$\log_{10}(m_{*,\mathrm{gross}}/\mathrm{M}_\odot) = m\log_{10}(M_{200\mathrm{c}}/\mathrm{M}_\odot) + Az + B. \tag{4.18}$$

The results in this work show that the most fitted values for the parameters m, A, and B are 1.64, 1/30, and -9.2 respectively for $10^8 < M_{200\mathrm{c}}/\mathrm{M}_\odot < 10^{11}$. Equation 4.18 well represents the relationship between $m_{*,\mathrm{gross}}$ and $M_{200\mathrm{c}}$ at all redshift. Although the galaxies are under the supernova feedback and reionisation suppression around $M_{200\mathrm{c}} \sim 10^8$ $\mathrm{M}_\odot$, the scatter of the plot is large enough to estimate the linear $\log - \log$ relation in the equation.

We note that, as the result from our model, $m_{*,\mathrm{gross}}$ is proportional to $M_{200\mathrm{c}}^{1.64}$, it implies that we have assumed the total number of ionising photons contributed by a halo of mass M_{halo}, $N_\gamma \propto M_{\mathrm{halo}}^{1.64}$ while the some works (e.g. Choudhury et al., 2009; Majumdar et al., 2014) assumed $N_{\gamma,\mathrm{total}} \propto M_{\mathrm{halo}}$. Some assumed $\dot{N}_\gamma \propto M_{\mathrm{halo}}$ (e.g. Iliev et al., 2006; Zahn et al., 2007; Dixon et al., 2016). The recent work in Mutch et al. (2015) also assumes $\dot{N}_\gamma$ within an ionised region to be proportional to $m_{*,\mathrm{gross}}/t_{\mathrm{H}}$, where t_{H} is the Hubble time, but does not explicitly show the relationship between $m_{*,\mathrm{gross}}$ and the halo mass.

4.8.5 Clustering of galaxies

To study the changes in galaxy clustering between the models, we use the 2-point correlation function (2PCF), $\xi(r)$, which describes the excess of probability dP over a random distribution of finding pairs of galaxies at a given separation r:

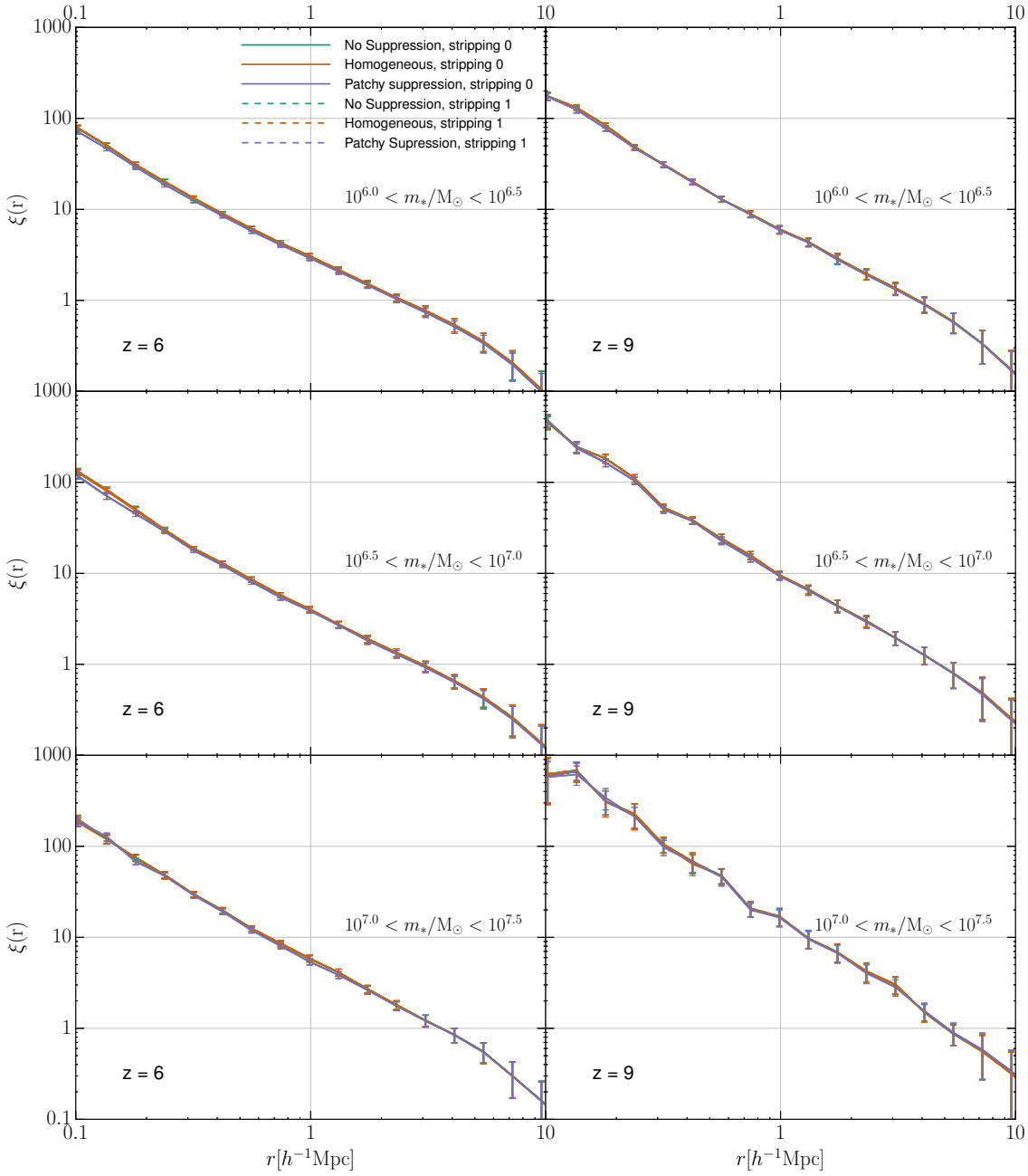

Figure 4.11: 2-point correlation functions as a function of comoving radius, r, of three subsets of galaxies within ionised regions from all models, $10^{6.0} < m_*/\mathrm{M}_\odot < 10^{6.5}$ (top), $10^{6.5} < m_*/\mathrm{M}_\odot < 10^{7.0}$ (middle), and $10^{7.0} < m_*/\mathrm{M}_\odot < 10^{7.5}$(bottom) at $z = 6$ (left panels) and $z = 9$ (right panels).

$$dP = n(\xi(r) + 1)dV, \tag{4.19}$$

where n is the number density of selected galaxies. We use a simple estimator described by the following formula:

$$\xi(r) = \frac{DD(r)}{RR(r)} - 1, \tag{4.20}$$

where $DD(r)$ is the number of data pairs separated a distance r between them, and $RR(r)$ is the number of random pairs at the same distance.

To calculate the errors of the 2PCF we use the Jack-Knife method described in Norberg et al. (2009). We divide the simulation box into 64 cubic subvolumes, and we measure the 2PCF 64 times excluding each time only one of the subsamples. We obtain the error from:

$$\Delta\xi(r) = \sqrt{\frac{N_{\rm JK} - 1}{N_{\rm JK}} \sum_{i=1}^{N_{\rm JK}} (\xi_i(r) - \bar{\xi}(r))^2}, \tag{4.21}$$

where $N_{\rm JK}$ is the number of Jack-Knife subsamples used and $\xi_i(r)$ corresponds to the measurement of $\xi(r)$ excluding the ith subsample.

Due to the suppression mechanism of the "Patchy" model which only suppress the haloes in ionised regions, the clustering of the galaxies selected by a short stellar mass range produced by this model should present some deviation from the reference the "No Suppression" model. Figure 4.11 shows the 2-point correlation functions as a function of comoving radius, r, of three subsets of galaxies in ionised regions from the models, $10^{6.0} < m_*/\mathrm{M}_\odot < 10^{6.5}$ (top), $10^{6.5} < m_*/\mathrm{M}_\odot < 10^{7.0}$ (middle), and $10^{7.0} < m_*/\mathrm{M}_\odot < 10^{7.5}$(bottom) at $z = 6$ (left panels) and $z = 9$ (right panels). Although the deviation is present, it is smaller than the errors calculated by Jack-Knife method. At $z = 9$, we notice the results from the "Patchy" model displays very slightly lower ξ, less than $\sim 10\%$, than other models at $r \sim 200$kpc with the selection of the galaxies with stellar mass of $10^6 - 10^7\mathrm{M}_\odot$. At $z = 6$, we notice less galaxy clustering from the selection of $10^{6.5} < m_*/\mathrm{M}_\odot < 10^7$ from the "Patchy" model. The 2PCF of the "Patchy" model shows that the probability to find other selected galaxies within comoving radius of $r = 200$kpc is about 20% lower than other models. For $10^6 < m_*/\mathrm{M}_\odot < 10^{6.5}$, the "Patchy" model also exhibits the 2PCF which is $< 10\%$ less than those estimated by the "No Suppression" and the "Homogeneous" models which appear to be identical.

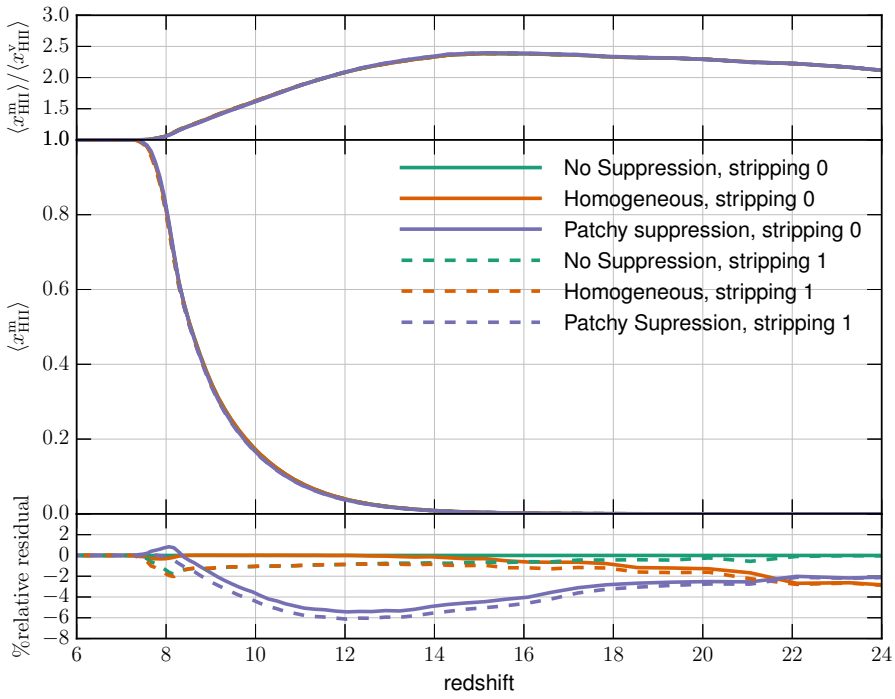

Figure 4.12: The time evolution of mass-weighted average H II fraction of the simulation box produced by each model (bottom) and corresponding ratios of mass-weighted and volume-weighted ionised fractions (top), which are equal to the mean density of the ionised regions in units of the mean density of the Universe. The bottom plot shows the percentage residual offset from the mass-weighted average H II fraction predicted by "No suppression" model using the stripping "0" scheme.

4.8.6 The effects on reionisation history

Figure 4.12 presents the time evolution of mass-weighted average H II fraction of the simulation box produced by each model (middle) and corresponding ratios of mass-weighted and volume-weighted ionised fractions (top), which are equal to the mean density of the ionised regions in units of the mean density of the Universe. The bottom plot shows the percentage residual offset from the mass-weighted average H II fraction predicted by "No suppression" model using the stripping "0" scheme. Since we assume the recombination is uniform, the mass-weighted H II fraction can be directly translated to the total number of ionising photons produced in the simulation. The plot shows that the suppression models in the stripping "1" scheme have about $1-2\%$ more suppression power than their suppression model in the stripping "0" scheme. This means the stripping "1" scheme which we propose to avoid the overestimated baryon contents in the galaxies makes only $1-2\%$ different in the total ionising photon emission. From $z \sim 15$, the reionisation histories of the "Homogeneous" and the "No Suppression" models are almost identical in both stripping schemes across the cosmic time, except very modest difference up to 0.2% at $z < 8$. Compared to "No Suppression" model, the "Patchy" model suppresses 6% of the global ionising photons at $z \sim 12$ then the suppression drops down to nothing at $z \sim 8.5$. After that, the "Patchy" model produce 1% and 2%, in the stripping "0" and "1" scheme respectively, more ionising photons than the "No Suppression" model at $z \sim 8$. This is because the "Patchy" model assumes that the infall into the low mass haloes in ionised regions is suppressed due to the ionising background. When they are accreted into larger systems (or even become ones themselves), their new hosts are able to use the baryons which have not been used in the previous time to supply new star formation. In the "No Suppression", this portion or baryons is accreted into those small haloes then cooled and condensed into galaxies. When these haloes are accreted into larger systems, their baryons remain locked up in the infalling satellite, making them unusable for the Type 0 galaxies to supply the star formation. Please note that there is no discrepancy in the ratio of mass-weighted and volume-weighted ionised fractions between suppression models and stripping schemes across the simulation time.

4.8.7 Ionisation morphology

In Figure 4.13, Figure 4.14, and Figure 4.15, we illustrate the evolution of H II fraction density fields of all models when their average mass-weighted ionised fractions, $\langle x_{\rm HII}^{\rm m} \rangle$ are 0.3, 0.5, and 0.7 respectively. In despite of having different suppression models, the ba-

sic features are similar for all models. At first, a large number of small, ~ 1 Mpc-size H II bubbles form. They are highly clustered, around the high number density regions of ionising sources. These small H II bubbles merge into nearby bubbles, and create larger ionised regions. When a half of the hydrogen in the simulation box is ionised, the percolation of H II regions can be spotted, and bubbles with size up to ~ 10 Mpc form. We note that these are perspective views of the ionisation field, and H II regions are not completely spherical, therefore they have different sizes depending on the viewing direction (Iliev et al., 2008b). The H II regions continue to grow. At $z \sim 8$, where $\langle x_{\rm HII}^{\rm m} \rangle \sim 0.7$, some extend to a few ~ 10 Mpc across.

In Figure 4.14, a trace of the baryon being restricted by the stripping "1" is shown in the partly ionised area near the centre of the plotted space. A circular light blue area appears to be 'cleaner' in the stripping "1" scheme (right panels) because the fluctuation in the mass history of the haloes near the region causes the extra star formation in the stripping "0" scheme (left panels). On the other hand, stripping "1" scheme has the mechanism to prevent the extra baryons from the mass fluctuation and avoid the extra star formation which we can see in stripping "0" scheme.

In Figure 4.15, an increasing of ionising photon emission in the "Patchy" model is seen in the lower left part of the plotted space where the model displays a larger ionised region in both stripping schemes. By using the same suppression model, the plot also shows that the ionised region is slightly larger in the stripping "0" scheme.

Since it is difficult to spot the differences in the 2D H II fraction plots and we want to extract more quantitative data, we continue to measure the size distribution of the ionised regions based on the spherical average method used in Zahn et al. (2007) and McQuinn et al. (2007). Figure 4.16 shows the probability distributions for the radius of ionised regions, R, at $\langle x_{\rm HII}^{\rm m} \rangle \sim 0.3$, 0.5, and 0.7 for all models. "No Suppression" and "Homogeneous" models show no discrepancy between them in all plots while the "Patchy" model predicts slightly less probability to find small ionised bubbles than the other two models. When $\langle x_{\rm HII}^{\rm m} \rangle \sim 0.7$, which is corresponding to $z \sim 8.1$, the "Patchy" model predicts, up to $\sim 10\%$ at $r = 0.1$Mpc, less probability to find small ionised bubbles with radius less than ~ 8Mpc than other models. The distributions reflect the suppression mechanism in the "Patchy" model which suppresses the baryonic infall of low mass haloes in the ionised regions. Therefore, the galaxies reside in small haloes which are normally responsible for the creation of small bubbles are suppressed and not able to create ionised bubbles on their own, hence there are less small bubbles. When they are merged into

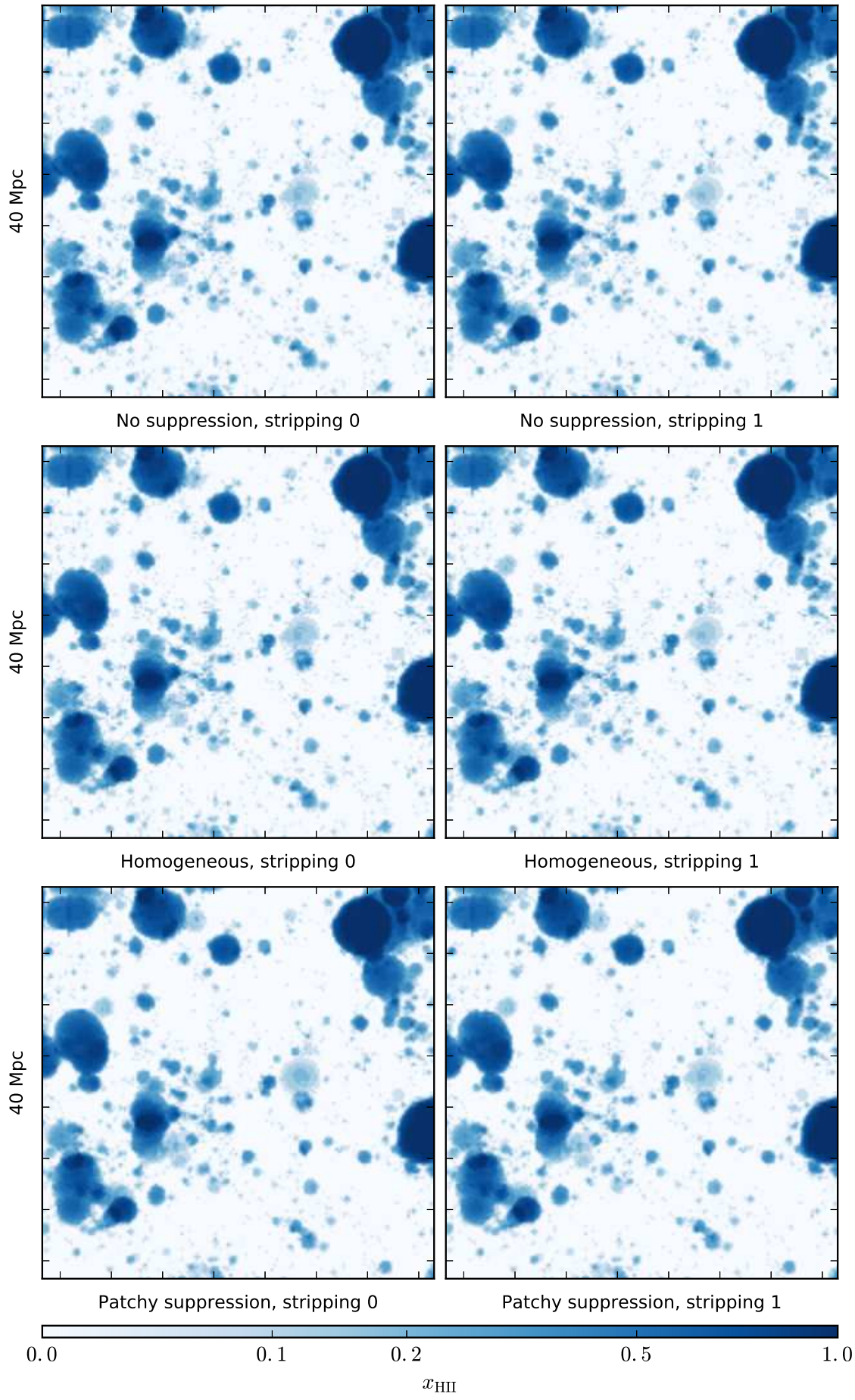

Figure 4.13: The ionization structure of a 3 Mpc thick slab at the redshift corresponding to a mass-weighted global neutral fraction of $\langle x_{\rm HII}^{\rm m} \rangle \sim 0.3$.

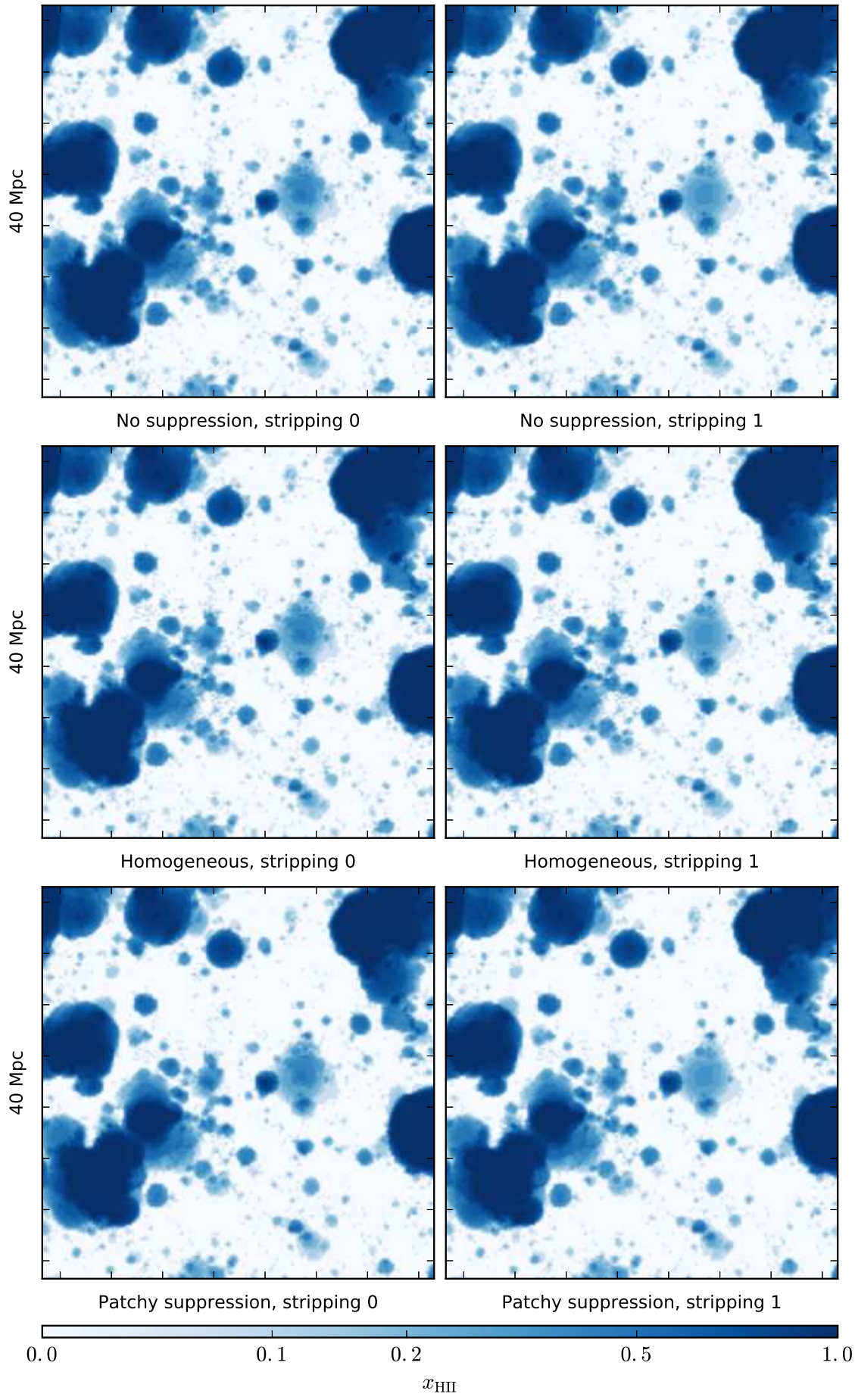

Figure 4.14: The ionization structure of a 3 Mpc thick slab at the redshift corresponding to a mass-weighted global neutral fraction of $\langle x_{\rm HII}^{\rm m} \rangle \sim 0.5$.

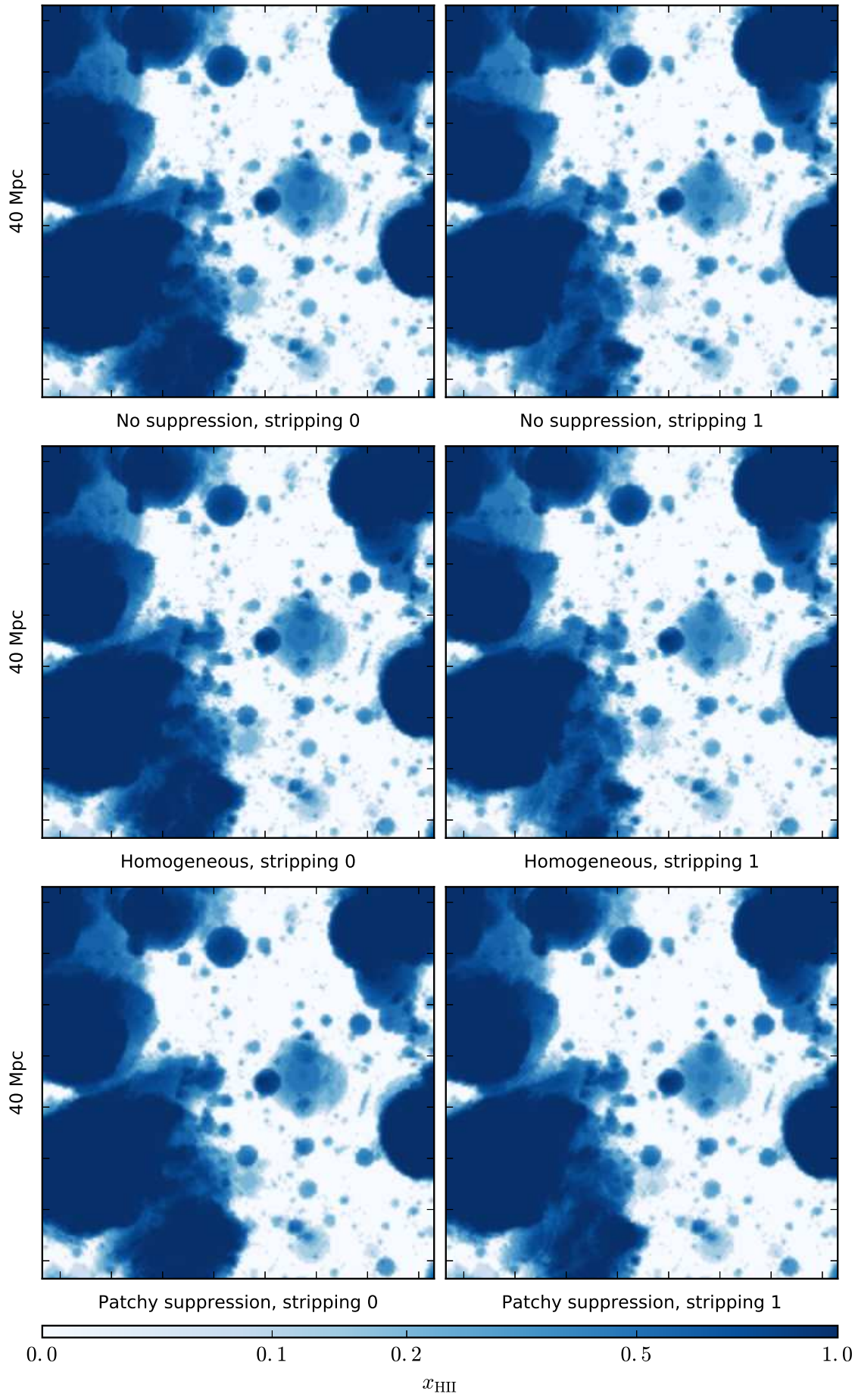

Figure 4.15: The ionization structure of a 3 Mpc thick slab at the redshift corresponding to a mass-weighted global neutral fraction of $\langle x_{\rm HII}^{\rm m} \rangle \sim 0.7$

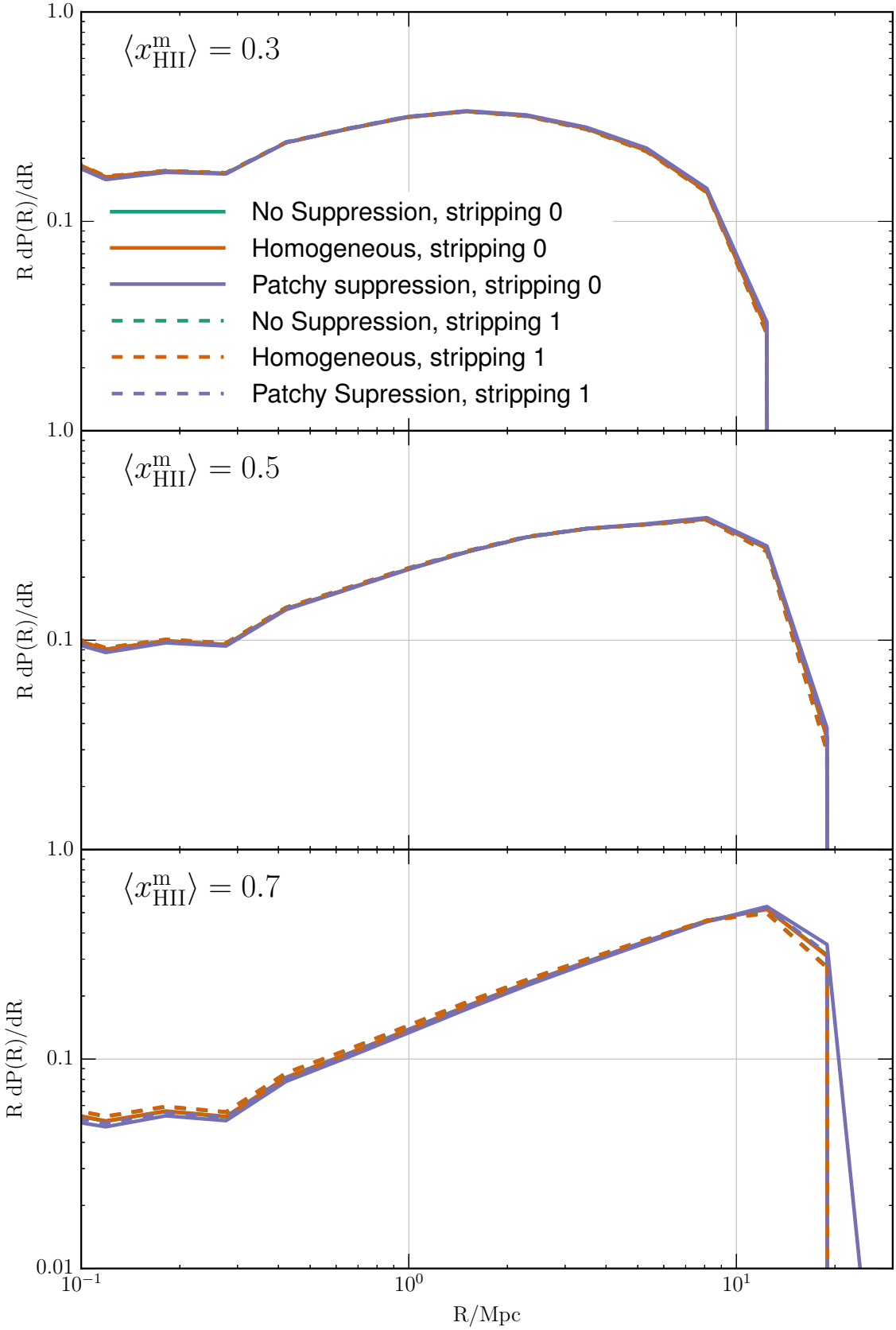

Figure 4.16: Size distributions of ionized regions at stages with averaged mass-weighted ionised fraction $\langle x_{\rm HII}^{\rm m} \rangle = 0.3$ (top), 0.5 (middle), and 0.7 (bottom).

larger systems, their unused baryons will be supplied to the host to form more new stars, hence there are more large bubbles. This effect might be caused by the difference in $\langle x_{\rm HII}^{\rm m} \rangle$ between each model representing $\langle x_{\rm HII}^{\rm m} \rangle \sim 0.7$ (see Section 4.8.6). However, it should not be the case since only 1% difference $\langle x_{\rm HII}^{\rm m} \rangle$ in can be seen between the "Patchy" model and others at this particular snapshot.

Chapter 5

Conclusions

In this work, we have presented a semi-analytic model of galaxy formation, using the latest version of L-Galaxies (Henriques et al., 2015), and fully coupled with a semi-numerical method of reionisation (Majumdar et al., 2014). It has been designed to study the formation and evolution of the first galaxies and their effects on the reionisation. By incorporating the method of reionisation simulation in the semi-analytic model, the precise stage of the surrounding IGM around each galaxy can be fed into the model of galaxy formation, thus the barynic infall can be modelled accordingly.

The model has been calibrated against the observed UV luminosity function at $z =$ 6, 7, and 8 from Bouwens et al. (2015b), and the 3-year *Planck* constraint on the free electron scattering optical depth (Planck Collaboration et al., 2015). The results in this work excellently agree with those observations and reproduce the star formation rate distribution function from Duncan et al. (2014) and the galaxy stellar mass function from Song et al. (2016) at high redshift.

The stripping "1" scheme was designed to avoid overestimation of the baryon contents inside galaxies caused by the fluctuation of halo mass across the cosmic history. While the stripping schemes show the effects in the baryon reservoirs inside galaxies especially in hot gas and ejected mass phases, it has a very minimal effect on the star formation. The discrepancy in the mass-weighted average H II fraction between the stripping schemes is only up to $\sim 2\%$ at $z \sim 8$ in this work.

We found no observationally detectable imprint in stellar masses of $m_* > 10^{5.5}$ at $z \geq 6$, while a recent similar work, Mutch et al. (2015), predicts the discrepancy to be observable in the stellar mass of $m_* < 10^{7.5}\mathrm{M}_\odot$. This might be due to the different prediction in the population of low mass galaxies. Our models estimate the number density of the galaxies with the stellar mass of $\sim 10^8\mathrm{M}_\odot$ at $z = 6$ which is about less than a half of that predicted

in Mutch et al. (2015).

Even by using the "Patchy" suppression model, we cannot find any detectable imprint in the stellar mass function with stellar mass of $m_* > 10^{5.5}\mathrm{M}_\odot$. However, the results from the 2PCF of galaxies with stellar mass between 10^6 and 10^7 $\mathrm{M}_\odot$ show the suppression of galaxy clustering due to the mechanism of the"Patchy" suppression. Thought it is small, this imprint could be directly detectable in a high redshift Lyman-α emitter (LAE) survey; e.g. *Subaru* telescope with its new Hyper Suprime-Cam, which is estimated to observe $\sim$5500 LAEs at $z = 6.6$ and $\sim$40 LAEs at z=7.3 (Ouchi 2012, private communication).

"Homogeneous" and "No Suppression" models agree very well in almost comparisons in this work, especially above $z = 8$ for which they are identical. However the result from Mutch et al. (2015) shows that the stellar mass of individual galaxies can vary by a factor of 2-3 between the suppression models.

Compared to L-Galaxies's standard, the "Homogeneous" model, the "Patchy" model is slightly more effective at suppressing the total ionising photon emission up to $z \sim 8.5$ which is corresponding to $\langle x_{\mathrm{HII}}^{\mathrm{m}} \rangle \sim 0.5$. When $z < 8$, the total ionising photon emission exceed those of the other models until the box in fully ionised at $z = 7.3$. However, the maximum difference of the total ionising photon emission between the models is only $5 - 6\%$ which occurred at $z \sim 15$.

Overall, we agree with Mutch et al. (2015) that the reionisation history of the inter galactic medium is insensitive to the reionisation feedback models that we use in this work. However, if the process of reionisation is more extended than is predicted by our model (for example, due to non-homogeneous recombinations in the IGM) then reionisation feedback may play a more effective role. Furthermore, the equilibrium cooling function for collisional processes from Sutherland and Dopita (1993) employed by L-Galaxies in the cooling model only depends on the metallicity and temperature of the gas but ignores radiative ionisation effects might cause L-Galaxies to overestimate the cooling rate of the gas which is exposed to the ionisation radiation.

Since the volume of the simulation in this work is too small to be used for studying the 21cm power spectrum (see Iliev et al., 2014), there is an ongoing work using this simulation method with a larger simulation with a slightly lower resolution, 244 h^{-1}Mpc with 4000^3 particles as described in Dixon et al. (2016), to investigate the effect of local UV suppression which could be detectable in the 21cm power spectrum measurements of current and upcoming radio surveys.

Acknowledgement

- We were supposed to use the simulation from Dixon et al. (2016) but there was some mistakes in the particle IDs of the simulation so we (my supervisor and I) needed to fix the code and reran the simulation using the same initial condition. Nevertheless, we stated in the text that the simulation was done by the referred work.

- The stripping "1" scheme was originally proposed by my colleague, Benoît Fournier, but it needed to be modified before it was used in this work.

Part III

Merger Tree Comparisons

Chapter 6

Sussing Merger Trees: The merger tree comparison project

6.1 Introduction

As we outlined the procedures of the semi-analytic models (SAMs) of galaxy formation in Part I and put an example into action in Part II of this thesis, we can see that they rely on the accuracy of both the individual halo catalogues themselves as well as the framework called merger tree that connects the halo catalogues from different snapshots together. Merger tree represents the history of haloes which can be tracked across the cosmic time by using a variety of tools and algorithms.

We will examine the accuracy of the trees (how often they link unrelated haloes together) and the smoothness of the tree growth. Both can lead to unrealistic galaxy growth within a SA model. The results presented in this part arise out of the Sussing Merger Tree workshop, that took place on July 7-12 2013. Different aspects relating to merger trees have been investigated. Currently, there are 3 publications from the workshop;

- Srisawat, C., Knebe, A., Pearce, F. R., Schneider, A., Thomas, P. A., Behroozi, P., Dolag, K., Elahi, P. J., Han, J., Helly, J., Jing, Y., Jung, I., Lee, J., Mao, Y.-Y., Onions, J., Rodriguez-Gomez, V., Tweed, D., and Yi, S. K. (2013). Sussing Merger Trees: The Merger Trees Comparison Project. MNRAS, 436:150162 (Srisawat et al., 2013).

- Avila, S., Knebe, A., Pearce, F. R., Schneider, A., Srisawat, C., Thomas, P. A., Behroozi, P., Elahi, P. J., Han, J., Mao, Y.-Y., Onions, J., Rodriguez-Gomez, V., and Tweed, D. (2014). SUSSING MERGER TREES: the influence of the halo finder.

MNRAS, 441:34883501 (Avila et al., 2014).

- Lee, J., Yi, S. K., Elahi, P. J., Thomas, P. A., Pearce, F. R., Behroozi, P., Han, J., Helly, J., Jung, I., Knebe, A., Mao, Y.-Y., Onions, J., Rodriguez-Gomez, V., Schneider, A., Srisawat, C., and Tweed, D. (2014). Sussing merger trees: the impact of halo merger trees on galaxy properties in a semi-analytic model. MNRAS, 445:41974210 (Lee et al., 2014).

The information in Part III will be very useful to introduce the necessary of proposing a new baryonic infall mechanism in 4.3.1. The contents will be organised as follows:

- Chapter 6 gives the introduction of the work, including terminology and brief descriptions of all merger tree generation algorithms used in the comparisons.
- In Chapter 7, we present the results from Srisawat et al. (2013), which give the overall results of the merger tree comparison project. This give a clear distinction between the results from different tree building methods, even their approaches are very similar. The mass fluctuations , especially during the merging processes, are discussed in detail.

6.2 Terminology

To avoid confusion, it is important that different researchers working on merger trees speak the same language. We define here the terminology used in this paper and would encourage others to adopt the same definitions:

- A **halo** is a dark-matter condensation as returned by a halo-finder (in our case AHF). For the purposes of other definitions below, we assume that the IDs of the particles attributed to each halo by the halo finder are known.
- Haloes may be spatially nested: in that case the outer halo is the **main halo** and the other haloes are **subhaloes**. Note that the assignment of main halos and subhaloes is a function of the halo-finder and one can envisage unusual geometries where this allocation is not obvious; nevertheless, the picture of subhaloes orbiting within larger ones ties in with our view of cosmic structure and is central to many SA models.
- If particles are allowed to be members of only one halo, (i.e. particles in sub-haloes are not included in the particle ID list of the main halo, and particles in overlapping haloes are assigned to just one of the two), then the haloes are said to be **exclusive**; otherwise they are **inclusive** (AHF falls into this latter category).

- Haloes are defined at distinct **snapshots**. Snapshots correspond to particular values of cosmic time and contain the particle IDs, mass, location & velocity for each dark matter particle in the simulation.

- For two snapshots at different times we refer to the older one (i.e. higher redshift) as A and the younger one (i.e. lower redshift) as B.

- A **graph** is a set of ordered halo pairs, (H_A, H_B), where H_A is older than H_B. It is the purpose of the merger-tree codes to produce a graph that best represents the growth of structure over cosmic time. H_A and H_B are usually taken from adjacent snapshots, but this is not a requirement as there are occasions where haloes lose their identity and then reappear at a later time.

- Recursively, H_A itself and progenitors of H_A are **progenitors** of H_B. Where it is necessary to distinguish H_A from earlier progenitors, we will use the term **direct progenitor**.

- Recursively, H_B itself and descendants of H_B are **descendants** of H_A. Where it is necessary to distinguish H_B from later descendants, we will use the term **direct descendant**.

- In this paper we are primarily concerned with **merger trees** for which there is precisely one direct descendant for every halo. Note that it is possible for haloes near the minimum mass limit to have zero descendants: we omit such haloes from our analysis.

- In the case that there are multiple direct progenitors, we require that precisely one of these be labelled the **main progenitor** – this will usually be the most massive, but other choices are permitted.

- The **main branch** of a halo is a complete list of main progenitors tracing back along its cosmic history.[1]

Over the course of writing this paper it became clear that there has been confusion in the past between what we call graphs and merger trees. Both are interesting in different contexts. We limit ourselves here to an investigation of merger trees which are the more relevant as an input to SA models.

[1] We note that, for main haloes rooted at $z = 0$, this main branch might more appropriately be called a trunk, but it seems unnecessary to introduce a new term for this specific purpose.

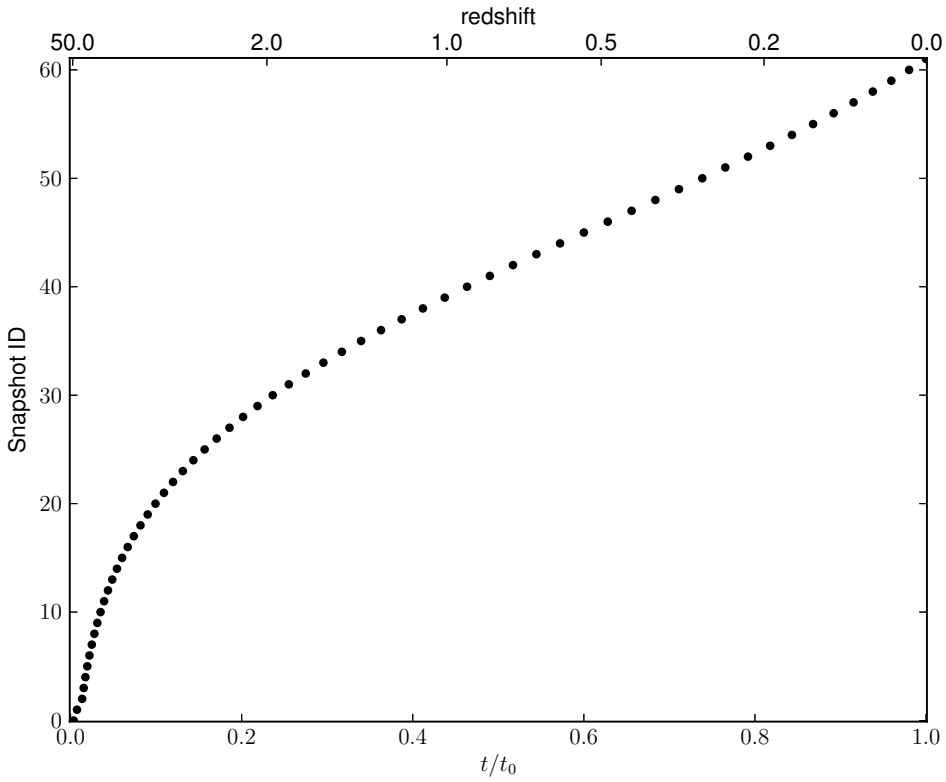

Figure 6.1: Snapshot ID versus time (lower x-axis, normalized to the present age of the Universe) and redshift (upper x-axis).

6.3 Input halo catalogues

The halo catalogues used for this paper are extracted from 62 snapshots of a cosmological dark matter only simulation undertaken using the GADGET-3 N-body code (Springel, 2005) with initial conditions drawn from the WMAP-7 cosmology (Komatsu et al., 2011b). We use 270^3 particles in a box of comoving width 62.5 h^{-1}Mpc, with a dark-matter particle mass of $9.31 \times 10^8 h^{-1}\mathrm{M}_\odot$. The snapshots are labelled 0, 1, 2, ..., 61 from redshift 50 to redshift 0, as indicated in Figure 6.1.

The main halo finder used in this paper is AHF[2] (Gill et al., 2004; Knollmann and Knebe, 2009). It locates local overdensities in an adaptively-smoothed density field as prospective halo centres. For each of these density peaks the gravitationally bound particles are determined. Only peaks with at least 20 bound particles are considered as haloes and retained for further analysis. The halo mass M_{200} is

$$M_{200} = 200\rho_c(z)\frac{4\pi}{3}R_{200}^3, \tag{6.1}$$

[2]The Amiga Halo Finder package is publicly available for download from `http://popia.ft.uam.es/AHF`

where $\rho_c(z)$ is the critical density of the Universe as a function of redshift z and R_{200} is the radius enclosing a mean density that equals 200 times the critical density.

AHF generates inclusive data sets (i.e. particles in subhaloes are also included in the main halo). As an input to the tree-building codes we provided the list of particle IDs associated with each halo, alongside information about the (kinetic plus potential) energy, position and velocity of each particle; we further made available the full halo catalogue containing, besides the usual mass, position, and bulk velocity, an abundance of additional information (e.g. energies, centre offsets, shapes, etc.).

The participants were asked to run their merger tree builders on the supplied data and return, for each halo, a list of progenitor haloes and (unless the halo was newly-created) the ID of a single main progenitor. For the purpose of comparing merger tree algorithms we restricted participants to use only the information described above and did not give them access to the raw N-body data. However, they were allowed to alter the original halo catalogues by adding extra "fake" haloes and removing some "unreliable" haloes where they felt that was appropriate.

6.4 Participated merger tree building algorithms

In this section we briefly describe, in alphabetical order, the participating merger tree codes. Further details of algorithms can be found in the accompanying references.

The participants were asked to build trees starting from our input halo catalogues described in Section 6.3. One of the features of a merger tree, as we define it, is that while an object can have multiple progenitors, only one descendant is allowed. But many of the algorithms tested did not, in the first instance, produce a tree. Instead they commonly built graphs that allowed multiple descendents of a single progenitor halo. To allow consistency and ensure a fair comparison we required each author to modify their algorithm to return a tree. Nevertheless, the central process of linking haloes together between snapshots remains and exploring the various ways of achieving this is the main purpose of this paper.

We note that some of the participating codes required modification in order to allow them to take as input the AHF halo catalogues that we used for this comparison project. To facilitate analysis of the returned merger trees, we have defined a common, minimal data output format (described in the Appendix), and this has also required minor modifications to some of them.

As a lot of methodology is similar across the various codes used here, we try to capture

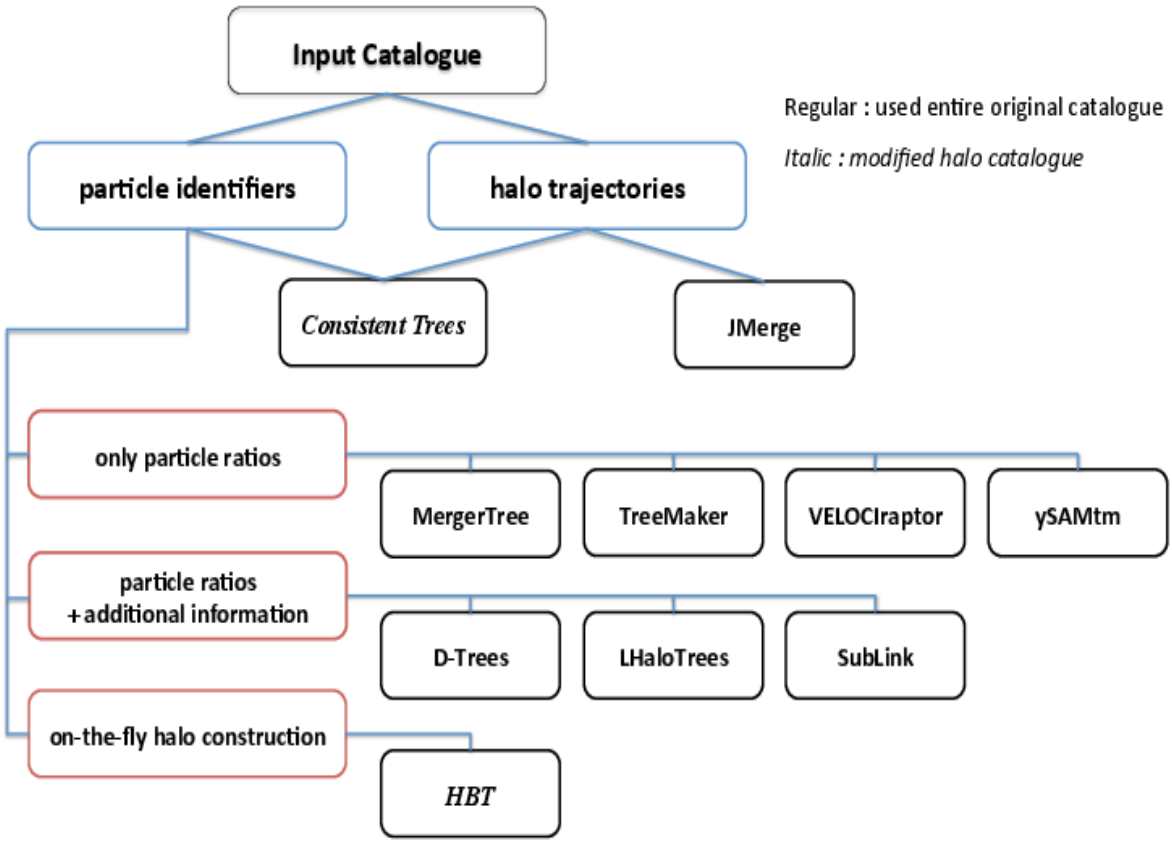

Figure 6.2: A summary of the main features and requirements of the different merger tree algorithms. For details see the individual descriptions in the text.

the main features and requirements in Figure 6.2 and Table 6.1. Note that only a single code doesn't use particle IDs to link haloes between snapshots: that potentially makes it more widely applicable to legacy data but leads to problems with misidentification of haloes, as will be seen later in Section 7.1 below.

Many tree-codes make use of a merit function

$$\mathcal{M}(H_A, H_B) = f(N_A, N_B, N_{A\cap B}), \tag{6.2}$$

where N_A and N_B are the number of particles in haloes H_A and H_B respectively. $N_{A\cap B}$ is the number of particles that are in both H_A and H_B, or

$$\mathcal{M}(H_A, H_B) = f(\mathcal{R}_{A\cap B}), \tag{6.3}$$

where $\mathcal{R}_{A\cap B}$ is the ranking (decreasing binding mass or increasng halocentric radius) of particles that are in both H_A and H_B. Such a function aims at identifying the most likely progenitor/descendant of a given halo. A few of them use additional information such as, for instance, the binding energy of the particles, properties of the haloes or information about the snapshot times.

6.4.1 Consistent Trees

The Consistent Trees algorithm (Behroozi et al., 2013b) first matches haloes between snapshots by identifying descendant haloes as those that have the maximum number of particles from a given progenitor halo. It then attempts to clean up this initial guess by simulating the gravitational bulk motion of the set of haloes given their known positions, velocities, and mass profiles as returned by the halo finder. From haloes in any given simulation snapshot, the expected positions and velocities of haloes at an earlier snapshot may be calculated. In some cases, obvious inconsistencies arise between the predicted and actual halo properties, such as missed satellite haloes (e.g. satellite haloes which pass too close to the centre of a larger halo to be detected) and spurious mass changes (e.g. satellite haloes which suddenly increase in mass due to temporary miss-assignment of particles from the central halo). These defects are repaired by substituting predicted halo properties instead of the properties returned by the halo finder. If a halo has no descendant a merger is assumed to have occurred with the halo exerting the strongest tidal field across it, unless no such suitable halo exists in which case the halo is presumed to have been spurious and this branch is pruned from the merger tree. This process helps to ensure accurate mass accretion histories and merger rates for satellite and central haloes; full

details of the algorithm as well as tests of the approach may be found in Behroozi et al. (2013b).

6.4.2 D-Trees

The D-Trees algorithm (Jiang et al., in preparation) is designed to work with the SUBFIND group finder, which (like AHF) can occasionally fail to detect haloes or subhaloes for one or more snapshots. It therefore allows for the possibility that descendants may be identified more than one snapshot later. Descendants are identified by following the most bound "core" of each group – i.e. those particles with the lowest total energy.

To find the descendant at snapshot B, of a group which exists at an earlier snapshot, A, the following method is used. For each group containing N_p particles the $N_{\rm link}$ most bound particles are identified, where $N_{\rm link}$ is given by

$$N_{\rm link} = \min(N_{\rm linkmax}, \max(f_{\rm trace} N_p, N_{\rm linkmin})) \tag{6.4}$$

with $N_{\rm linkmin} = 10$, $N_{\rm linkmax} = 100$ and $f_{\rm trace} = 0.1$. Descendant candidates are those groups at snapshot B that received at least one of the $N_{\rm link}$ most bound particles from the earlier group. If any of the descendant candidates received a larger fraction of their $N_{\rm link}$ most bound particles from the progenitor group than from any other group, then the descendant is chosen from these candidates only and the group at snapshot A will be designated the main progenitor of the chosen descendant; otherwise all candidates are considered. The descendant of the group at snapshot A is taken to be the remaining candidate which received the largest fraction of the $N_{\rm link}$ most bound particles of the progenitor group. For each group at snapshot B, this method identifies zero or more progenitors of which at most one may be a main progenitor. Note that it is not guaranteed that a main progenitor will be found for every group.

If a group is not found to be the main progenitor of its descendant, this may indicate that the group has merged with another group and no longer exists in the simulation. However, it is also possible that the group finder has simply failed to identify the object at the later snapshot. In order to distinguish between these cases it is necessary to search multiple snapshots.

For each snapshot A in the simulation descendants are identified at later snapshots in the range $A + 1$ to $A + N_{\rm step}$ using the method described above. For each group at snapshot A this gives up to $N_{\rm step}$ possible descendants. One of these descendants is picked for use in the merger trees as follows: if the group at snapshot A is the main progenitor of one or more of the descendants, the earliest of these descendants that does not have a

main progenitor at a snapshot later than A is chosen. If no such descendant exists, the earliest descendant found is chosen irrespective of main progenitor status.

This results in the identification of a single descendant for each group, which may be up to $N_{\rm step}$ snapshots later. Each group may also have up to one main progenitor which may be up to $N_{\rm step}$ snapshots earlier.

6.4.3 HBT

The Hierarchical Bound Tracing (HBT) algorithm (Han et al., 2012b) is a tracking halo finder in the sense that it uses information from earlier snapshots to help derive the latest halo catalogue. As such it naturally builds a merger tree. Starting from high redshift, main haloes are identified as they form. The particles contained within these haloes are then followed explicitly through subsequent snapshots, generating a merger tree down to main halo level at the first stage. To extend the merger tree down to subhalo level, HBT continues the tracing of merged branches, identifying the set of self-bound particles that remain for every progenitor halo. These self-bound remnants are defined as descendant haloes of their progenitors. With this kind of tracking, each halo has at most one progenitor, which defines its main branch. The main branch extends until the number of particles contained in the bound halo remnant drops below 20 particles. When this occurs a final tracking step is undertaken to determine which halo it has fallen into, adding minor branches to the tree.

The major challenge in this method is to robustly track haloes over long periods, and HBT has been specifically tuned to achieve this. In addition, the merging hierarchy among progenitor haloes is utilized to efficiently allow satellite-satellite mergers or satellite accretion inside satellite systems.

Note that HBT is not designed to be a general purpose treebuilder for external halo catalogues. To generate the trees used in this paper, HBT was run using only the main haloes from the supplied catalogue as described in Section 6.3 as input. It then outputs its own list of haloes and calculates the relevant properties for them, as well as returning the merger tree built on top of these haloes.

HBT outputs exclusive halos. In order to give a mass which matches that of AHF halos as closely as possible, for each halo, we first calculate an 'exclusive' mass according to Equation 6.1 using only particles from the halo itself. Then we add to each halo the exclusive mass of all its subhaloes, to give an 'inclusive' mass, which we use throughout this paper.

6.4.4 JMerge

The JMerge algorithm constructs a merger tree purely from aggregate properties (the position, centre-of-mass velocity and mass) of the haloes identified by a halo finder (i.e. it does not require the individual particle positions or particle IDs). It compares halo catalogues from two snapshots separated by a known time interval. For the two sets of haloes at times A and B, a new position is calculated for the centre of each halo by moving the A haloes forward in time by half the timestep, and the B haloes backwards by half the timestep assuming that they are moving at constant velocities. Then, starting from the most massive halo and working towards smaller masses, for each halo in A, a best match on position is found to a halo in B, together with constraints on the allowed change in mass and maximum circular velocity. Mass is allowed to shrink by a factor of up to 0.7, and to grow by a factor of up to 4. The search distance is limited to twice the radius at which the enclosed density is 200 times the background density plus four times the distance the halo has moved during the timestep. At this stage, each halo in B can only be claimed once. This process attempts to trace haloes growing over time.

For those haloes that do not find an unclaimed descendant in B, two other processes are implemented. Firstly, mergers are accounted for by finding so far unmatched haloes at time A that can accrete onto B targets already accounted for, whilst still limiting the total mass of the direct progenitors of each descendant to less than $1/0.7$ times its mass. Secondly, haloes that cannot find a suitable match are deemed to be numerical artifacts and are pruned from the tree.

6.4.5 LHaloTree

L-HaloTree was the first merger-tree algorithm to construct trees based on subhaloes instead of main halos. The LHaloTree algorithm is described in the supplementary information of Springel et al. (2005c) and the reader is referred there for further details. In short, to determine the appropriate descendant, the unique IDs that label each particle are tracked between outputs. For a given halo, the algorithm finds all halos in the subsequent output that contain some of its particles. These are then counted in a weighted fashion, giving higher weight to particles that are more tightly bound in the halo under consideration, as listed in Table 6.1, and the one with the highest count is selected as the descendant. In this way, preference is given to tracking the fate of the inner parts of a structure, which may survive for a long time upon infall into a bigger halo, even though much of the mass in the outer parts can be quickly stripped.

To allow for the possibility that halos may temporarily disappear for one snapshot, the process is repeated for Snapshot n to Snapshot $n+2$. If either there is a descendant found in Snapshot $n+2$ but none found in Snapshot $n+1$, or, if the descendant in Snapshot $n+1$ has several direct progenitors and the descendant in Snapshot $n+2$ has only one, then a link is made that skips the intervening snapshot.

6.4.6 MergerTree

The MergerTree routine forms part of the publicly available Amiga halo finder (AHF) package. It is a simple particle correlator: it takes two particle ID lists (ideally coming from an AHF analysis) and identifies for each object in list B those objects in list A (at the previous snapshot) with which there N or more particles in common ($N = 10$ for this comparison). Despite its name, therefore, it produces a graph mapping the connections between objects rather than a tree, as each halo can have multiple descendants.

MergerTree also identifies a unique main progenitor for each object in list B as found in list A. It achieves this by maximising a merit function (as shown in Table 6.1) This has proven extremely successful (Klimentowski et al., 2010; Libeskind et al., 2011; Knebe et al., 2013a). The code can hence not only be used to trace a particular object backwards in time (or forward, depending on the temporal ordering of files A and B), but also to cross-correlate different simulations (e.g. different cosmological models run with the same phases for the initial conditions).

To create an actual tree, we need to ensure that each halo has a unique descendant. This is guaranteed by running MergerTree in a novel mode that applies the same merit function in both directions when correlating two files. In practice this links haloes that share the largest fraction of particles between the two snapshots as well as forcing a choice between multiple possible descendants (of which now only the one maximising the merit function in the direction $A \mapsto B$ is kept). The use of a merit function also eliminates any need for all the particles in the input halo catalogues to only belong to a single object: $\mathcal{M}_{A_i B_j}$ automatically takes care of particles that have been assigned to multiple objects.

6.4.7 SubLink

SubLink (Rodriguez-Gomez et al., in prep.) constructs merger trees at the subhalo level. A unique descendant is assigned to each subhalo in three steps. First, descendant candidates are identified for each subhalo as those subhaloes in the following snapshot that have common particles with the subhalo in question. Second, each of the descendant candid-

ates is given a merit function specified in Table 6.1. Third, the unique descendant of the subhalo in question is the descendant candidate with the highest merit function.

Sometimes the halo finder does not detect a small subhalo that is passing through a larger structure, because the density contrast is not high enough. SubLink deals with this issue in the following way. For each subhalo from snapshot S_n, a 'skipped descendant' is identified at S_{n+2}, which is then compared to the 'descendant of the descendant" at the same snapshot. If the two possible descendants at S_{n+2} are not the same object, we keep the one obtained by skipping a snapshot since, by definition, it has the largest score at S_{n+2}.

Once all descendant connections have been made, the main progenitor of each subhalo is defined as the one with the 'most massive history' behind it, following De Lucia & Blaizot (2007). This information is rearranged into fully-independent merger trees.

6.4.8 TreeMaker

The TreeMaker algorithm was developed for the SA model `GalICS` (Galaxies in Cosmological Simulations) (Hatton et al., 2003). It was first used on Friends-of-Friends haloes (Davis et al., 1985b), and later applied to main haloes and subhaloes extracted from a cosmological simulation with the `AdaptaHOP` group finder (Aubert et al., 2004; Tweed et al., 2009). The code associates haloes from two consecutive time steps, listing all progenitors (including particles accreted from the background) and descendants (multiple descendants being allowed even if particles lost to the background are ignored). Here "background" refers to particles not in any halo at the current time. This first step is completed by using the particle IDs as tracers to identify haloes. Under our scheme a particle can only belong to one single halo at a given step, meaning a particle in a subhalo belongs only to that subhalo and not to any enclosing halo.

In order to create a "usable" merger tree a simplification stage is required. Exactly one descendant per halo is selected and the list of progenitors updated to reflect this selection. Selecting this unique descendant requires the use of a merit function. The first versions of TreeMaker used a shared merit function. For this study, we tested various modifications of this selection, but all gave similar results. We therefore include in this paper only the normalised merit function $\mathcal{M}_1$ as shown in Table 6.1.

6.4.9 VELOCIraptor

The halo merger tree algorithm used in VELOCIraptor is based on a particle correlator: that is the algorithm compares two (or more) *exclusive* particle ID lists and produces a catalogue of matches for each object in each list. Specifically, for each object i in catalogue A, the algorithm finds all objects j in catalogue B that share particles, and calculates the strength of each connection using the merit function $\mathcal{M}_1$ as shown in Table 6.1. The search for connections is done in both directions. Any connection with a merit function within Poisson fluctuations, $\mathcal{M}_{A_iB_j} \leq 1/(N_{A_i}N_{B_j})$, is ignored. The connection that maximises $\mathcal{M}$ for $A \rightarrow B$ is deemed the unique descendant (note that the original code returned a graph that did not enforce this requirement of uniqueness). This approach is used as particle ID lists produced by VELOCIraptor contain not only particles belonging to bound (sub)haloes but also those in physically diffuse tidal debris. Consequently, tracking object centres or weighting particles by a measure of how bound they are is meaningless. Note that tidal debris candidates, due to their physically diffuse nature, can be artificially fragmented into several VELOCIraptor groups. For example, a single bound (sub)halo identified at time A is found to be the progenitor of several tidal debris fragments at time B. Matching $B \rightarrow A$, the fragments identify the (sub)halo as the primary progenitor, however, the (sub)halo will identify the largest tidal fragment as its primary descendant. For proposes the of a this paper, the other fragments are ignored. However, in the general merger graph produced by VELOCIraptor, these fragments are flagged as secondary descendants if fragment shares $\geq 5\%$ of particles with the primary progenitor.

6.4.10 ySAMtm

The tree-making algorithm ySAMtm (Jung et al., in preparation) was developed to build dark matter halo merger trees for the semi-analytic model `ySAM` (Lee and Yi, 2013). It uses the particle information from two snapshot files or the particle IDs and locations from a pre-calculated halo catalogue. First the 'shared mass', the mass contribution of all progenitor haloes to each descendant halo, is calculated. At this stage, particles are matched between haloes in the two snapshots by using the particle IDs. Individual particles are only included in a single halo or subhalo and are not listed as members of the host halo of the subhalo. Secondly, in order to convert our graph into an actual tree that could be used by semi-analytic models, we define a unique descendant halo of each progenitor halo by determining which descendant halo has the most shared mass among all descendants of the progenitor halo, unless there exists a smaller halo which receives a larger fraction of its

mass from the same progenitor. In this case we determine that the smaller one is the most likely descendant halo of the progenitor even if its shared mass is not the largest amongst all the descendants. This avoids defining the smaller descendant halo as a newly-formed halo when it contains many particles that were members of an existing halo in the previous snapshot. This process creates a true tree where one descendant halo can have multiple progenitor haloes, while each progenitor halo has a unique descendant halo. Among those progenitors, the main progenitor is determined by maximising the merit function $\mathcal{M}_2$ in Table 6.1.

Table 6.1: A summary of the features and requirements of merger tree algorithms (for details see individual descriptions in the text). Columns: (i) Code name; (ii) Particle properties used to produce the merger trees; (iii) AHF halo properties used to produce the merger trees (M_{200}-mass, **r**-position, **v**-velocity, $V_{\rm max}$-maximum rotation speed of the halo); (iv) the merit function used to estimate descandants; (v) the merit function used to estimate the main progenitor; (vi) the number of consecutive snapshots used to determine descendants/progenitors at each snapshot. $\mathcal{M}_1 = N^2_{A\cap B}/(N_A N_B)$, $\mathcal{M}_2 = N_{A\cap B}/N_B$, $\mathcal{M}_3 = N_{A\cap B}$, $\mathcal{M}_4 = \sum_j \mathcal{R}^{-2/3}_{(A\cap B)_j}$, $\mathcal{M}_5 = N_{A\cap B}/N_B$ for most bound particles only.

	Particle properties used	AHF halo properties used	D.Merit Func.	P.Merit Func.	#Snapshots used
Consistent Trees*	PID	M_{200}, **r**, **v**, $V_{\rm max}$	$\mathcal{M}_3$	Trajectory Est.	4***
D-Trees	PID,binding energy	—	$\mathcal{M}_5$	$\mathcal{M}_5$	5***
HBT*	PID,position,velocity	—	—	—	2
JMerge	—	M_{200}, **r**, **v**, $V_{\rm max}$	Trajectory Est.	Trajectory Est.	2
LHaloTree	PID,binding energy**	—	$\mathcal{M}_4$	Most massive halo	3
MergerTree^i	PID	—	$\mathcal{M}_1$	$\mathcal{M}_1$	2
SubLink	PID,binding energy**	—	${\mathcal{M}_4}^3$	Most massive history	3
TreeMaker	PID	—	$\mathcal{M}_1$	$\mathcal{M}_1$	2
VELOCIraptor	PID	—	$\mathcal{M}_1$	$\mathcal{M}_1$	2
ySAMtm	PID	—	$\mathcal{M}_2$	$\mathcal{M}_2$	2
*modify catalogue ^{i} uses the inclusive particle convention	**use the distance from halo's centre for this comparison				***Users specify but these numbers are used for this comparison

Chapter 7

General Comparisons

7.1 Tree structure

In this section we look at the structure/geometry of trees. This includes a comparison of measurable quantities like the tree-length along the main branch, the tree-branching at every step, and the general consistency of the tree (i.e. possible mis-identification of descendants).

7.1.1 Length of main branches

The most basic requirement of a tree-building code is to trace haloes back in time. The length of the main branch gives a measure of how long single haloes can be followed through the complicated merger history of structure formation. Figure 7.1 shows the number, N, of $z = 0$ haloes that have main branches extending for a given number of snapshots, l, for all haloes within three different mass-ranges: haloes with $M_{200} < 10^{11}\,h^{-1}\mathrm{M}_{\odot}$ (less than ~ 100 particles) are shown in the top panel, $2 \times 10^{11}\,h^{-1}\mathrm{M}_{\odot} < M_{200} < 5 \times 10^{11}\,h^{-1}\mathrm{M}_{\odot}$ in the middle panel, and $M_{200} > 10^{12}\,h^{-1}\mathrm{M}_{\odot}$ (more than ~ 1000 particles) in the bottom panel.

Large haloes (bottom panel of Figure 7.1) tend to have long main branches with $l = 30$–50, which is in agreement with the picture of bottom-up structure formation, where larger objects form through repeated mergers of smaller ones.

As one moves to smaller haloes the proportion of short branches increases. For $M_{200} < 10^{11}\,h^{-1}\mathrm{M}_{\odot}$ the number of main branches per length is roughly constant from $l = 0$ until about $l = 30$ (corresponding to $z \approx 5$) and only drops to zero beyond $l \approx 50$ ($z \approx 10$). Thus, even in a hierarchical structure formation scenario, dwarf-sized haloes *that survive to the current day* have a wide variety of formation times.

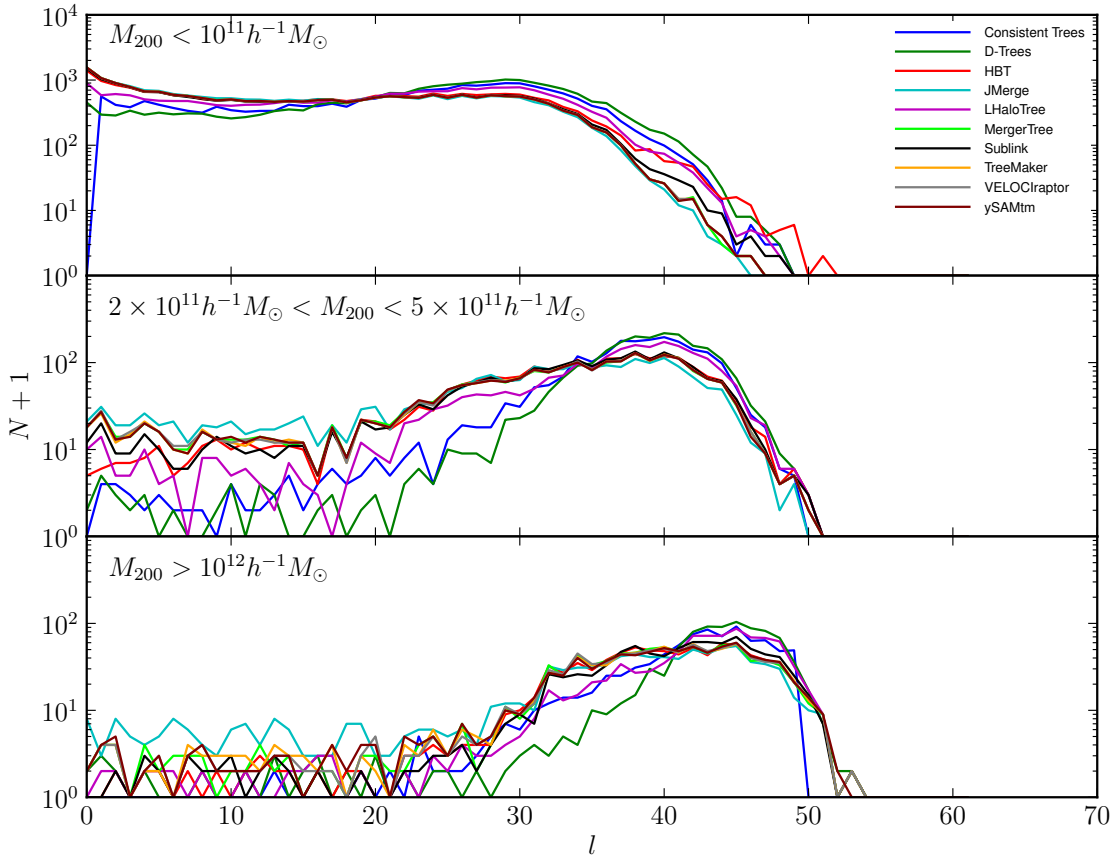

Figure 7.1: The length of the main branch for haloes identified at $z = 0$ (Snapshot 61). The ordinate is $l = 61 - S$, where S is the snapshot number at the high-redshift end of the main branch. The upper, middle and lower panels show the halo mass ranges at $z = 0$, as indicated in the panel, which correspond to roughly < 100, 200-500 and > 1000 particles respectively.

One oddity in Figure 7.1 is that most of the tree codes find a few large haloes with very short main branches which is in contradiction to the common picture of structure formation. Further investigation of these branches show that they are either truncated due to a non-identification by the halo finder, or are due to an error in the halo assignment of the tree building codes.

One such example is pictured in Figure 7.2 which shows two similarly-sized haloes merging almost head-on. The red and blue circles show the two haloes at $z = 0$ (right-hand column) and then traced back in time over several snapshots (successive columns to the left - note that we have chose to omit Snapshot 58 as it added little to the plot). The AHF halo finder (and other halo finders behave in a similar manner) assigns most of the mass in overlapping objects to a single object, treating the other as substructure. Unfortunately, this assignment can change between snapshots so that haloes centred on the same clump of highly-bound particles can fluctuate wildly in size. Different tree codes handle this in different ways, illustrated in the different rows of Figure 7.2.

- MergerTree fails to find a match for the smaller of the two haloes at Snapshot 60 and does not seek a match at earlier times. This halo therefore has no links in its merger tree and appears to be created intact in the final snapshot. The other merit function codes that use just 2 snapshots (TreeMaker, VELOCIraptor and ySAMtm) behave in the same manner, as, in this case, does JMerge.

- LHaloTree does something similar, but due to its use of the most-bound particles for halo identification, it matches the smaller of the two haloes at $z = 0$ to the large one from the previous snapshot. While LHaloTree can cross-match haloesby skipping a snapshot, that isn't applied here as a descendent halo exists.

- D-Trees does the same as LHaloTree on Snapshot 60, but also manages to link together the larger of the two haloes between Snapshots 61 & 59. This results in a fluctuating mass for the both haloes, (low-high-low for red, high-low-high for blue).

- SubLink also manages to cross-match the larger of the haloes between Snapshots 61 & 59 but chooses a different association for the halo in Snapshot 60, thus avoiding the large mass fluctuation. It links the smaller of the two halos in Snapshot 61 directly to that in Snapshot 59, skipping over the intermediate snapshot.

- Consistent Trees goes one step further and introduces a fake halo in Snapshot 60 to avoid a link in the merger tree that extends over more than one snapshot.

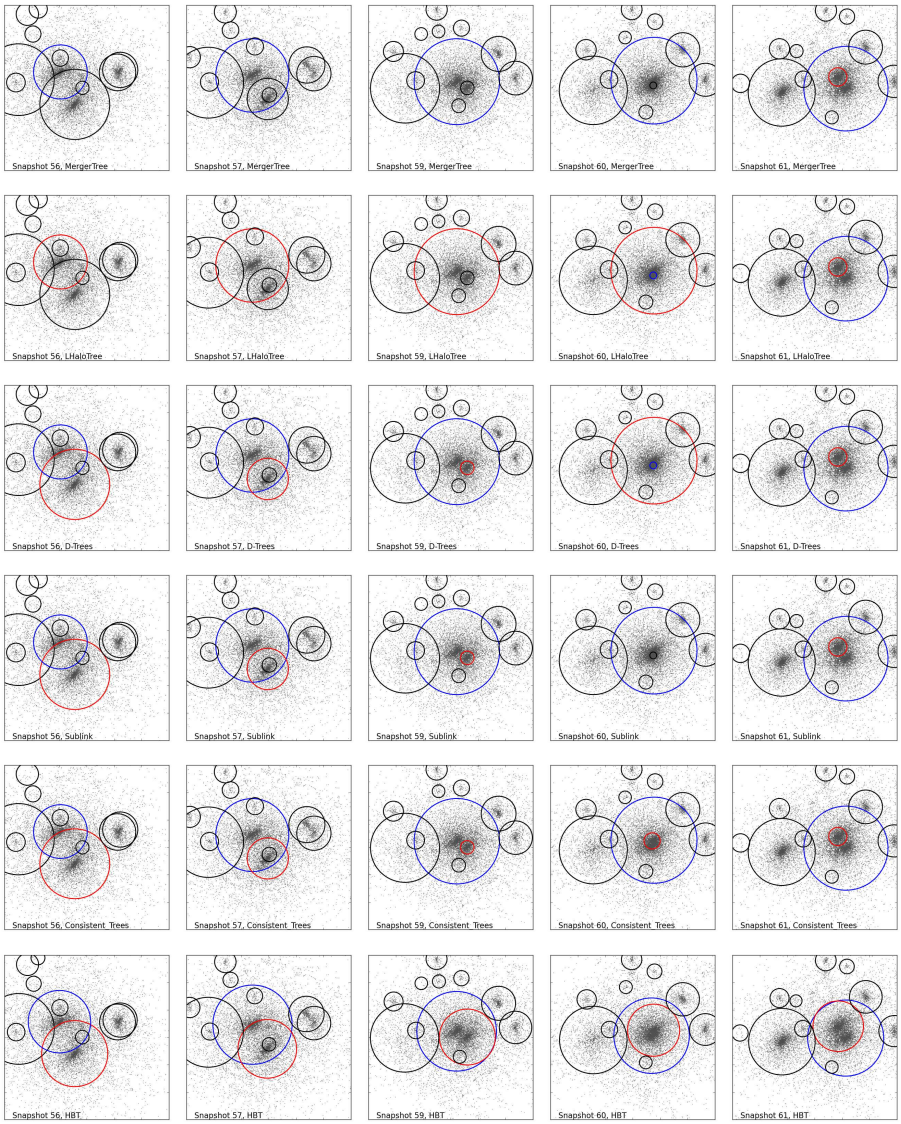

Figure 7.2: An example of the merger of two haloes where the fluctuating centering and size of causes difficulties for the merger-tree algorithms. The red and blue circles show two haloes selected at $z = 0$ (right-hand column) and then traced back in time over several snapshots (successive columns to the left - note that we have chosen to omit Snapshot 58 as it added little to the plot).

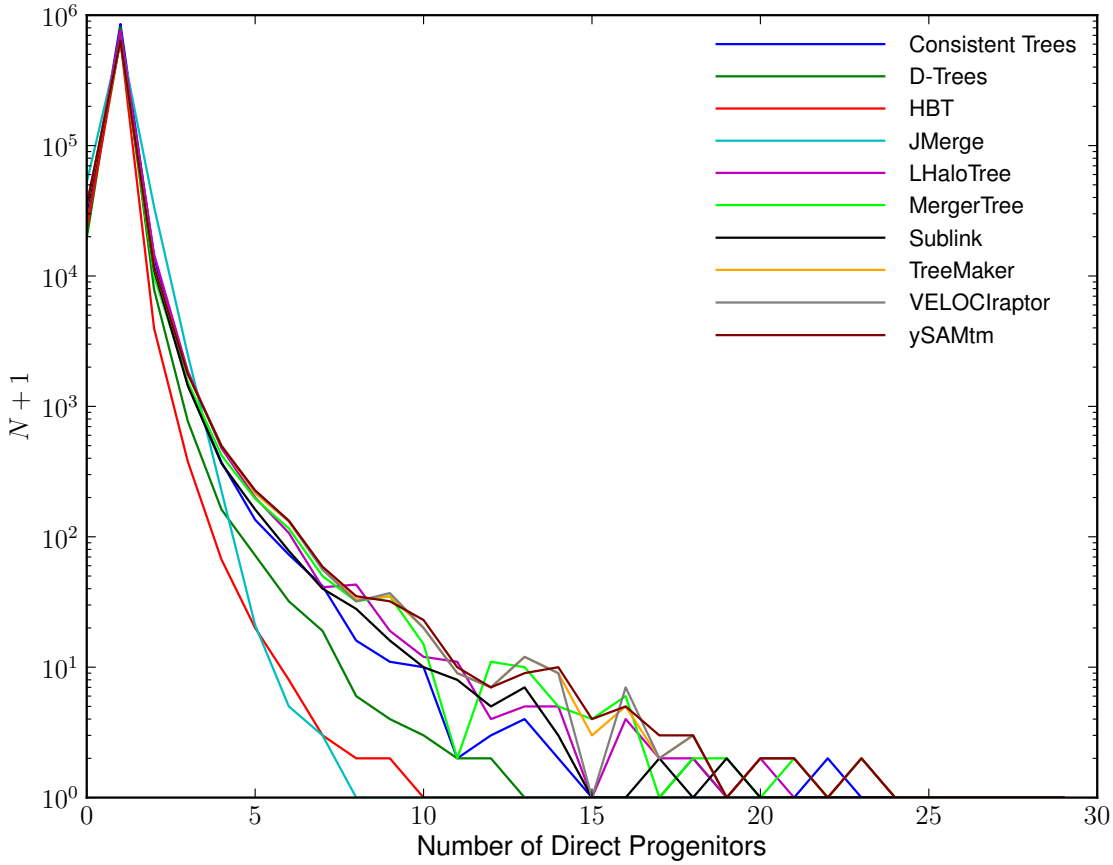

Figure 7.3: Histograms of the number of haloes with N_{dprog} direct progenitors, using all halos from $z = 0$ to $z = 2$.

- Finally, HBT redefines both haloes and outputs a smoother variation of mass over time.

From these descriptions, it may seem like the above is an ordered list of improving performance, from top to bottom. However, we stress that this is true only for this particular merging event and that different codes cope better in different situations. The purpose here was more to illustrate the variety of behaviours that are possible.

7.1.2 Branching ratio

Another interesting statistical quantity is the number of branches (i.e. the number of direct progenitors) at every node of the merger tree. This will depend upon the spacing between snapshots, and so the precise values are not important, but the differences between algorithms are still of interest.

In Figure 7.3 we plot the number of tree nodes with N_{dprog} direct progenitors, including all haloes between redshift zero and two. In this range the timestep Δt between snapshots

is roughly constant with $\Delta t \sim 0.4\,\mathrm{Gyr}$. The most common situation is to have a single progenitor (i.e. the halo existed in the previous snapshot but no merging took place), followed by zero progenitors (i.e. the halo appears for the first time). However, in some cases, and depending on the tree-builder, the number of direct progenitors can exceed 20.

HBT has the lowest branching ratio, perhaps because it allows itself to modify the halo catalogue to extend the life of subhaloes. JMerge also has a low branching number because its non-use of particle IDs gives it freedom to link together haloes that other algorithms classify as unrelated. Next come D-Trees and Consistent Trees which both use information extended over several timesteps to follow haloes that temporarily disappear (for instance when a subhalo comes close to the centre of its host halo).

Although multiple direct progenitors are rare, it can be seen that the choice of tree code can make a significant difference to the ability to follow substructures and hence to the length of time a subhalo exists before it is subsumed into the host halo.

7.1.3 Misidentifications

Most tree-building algorithms link together haloes on the basis of having particles in common. However, there are some that do not (in this paper, JMerge), and there are occasions when this association is not clear-cut. So we wish to test how often an obvious mis-identification occurs.

One way of doing this is to quantify how far haloes are displaced from their expected locations in moving from one snapshot to the next. This is hard to predict for sub-haloes that may be moving around inside a larger object and so we restrict our attention to main haloes only. To measure this deviation we use the statistic

$$\Delta_r = \frac{|\mathbf{r}_B - \mathbf{r}_A - 0.5(\mathbf{v}_A + \mathbf{v}_B)\,(t_B - t_A)|}{0.5\,(R_{200A} + R_{200B} + |\mathbf{v_A} + \mathbf{v_B}|(t_B - t_A))} \tag{7.1}$$

which stays small as long as there is approximately uniform acceleration and no error in the halo linking. Here t is cosmic time, $\mathbf{r}$ & $\mathbf{v}$ are the haloes' positions and velocities, and R_{200} the radius that encloses an overdensity of 200 times the critical density. The subscripts A and B refer to two linked haloes along the main branch of any tree.

Figure 7.4 shows a histogram of Δ_r for each algorithm, for all main haloes and their corresponding main progenitors. Most algorithms agree on the bulk of the distribution, and this likely represents the true behaviour for the AHF haloes considered here, with deviations from $\Delta_r = 0$ being caused by curved trajectories and/or merging of subhaloes. The difference in HBT's result from the others is partly due to different tree-links but also because the HBT halo catalogue has an intrinsically lower Δ_r.

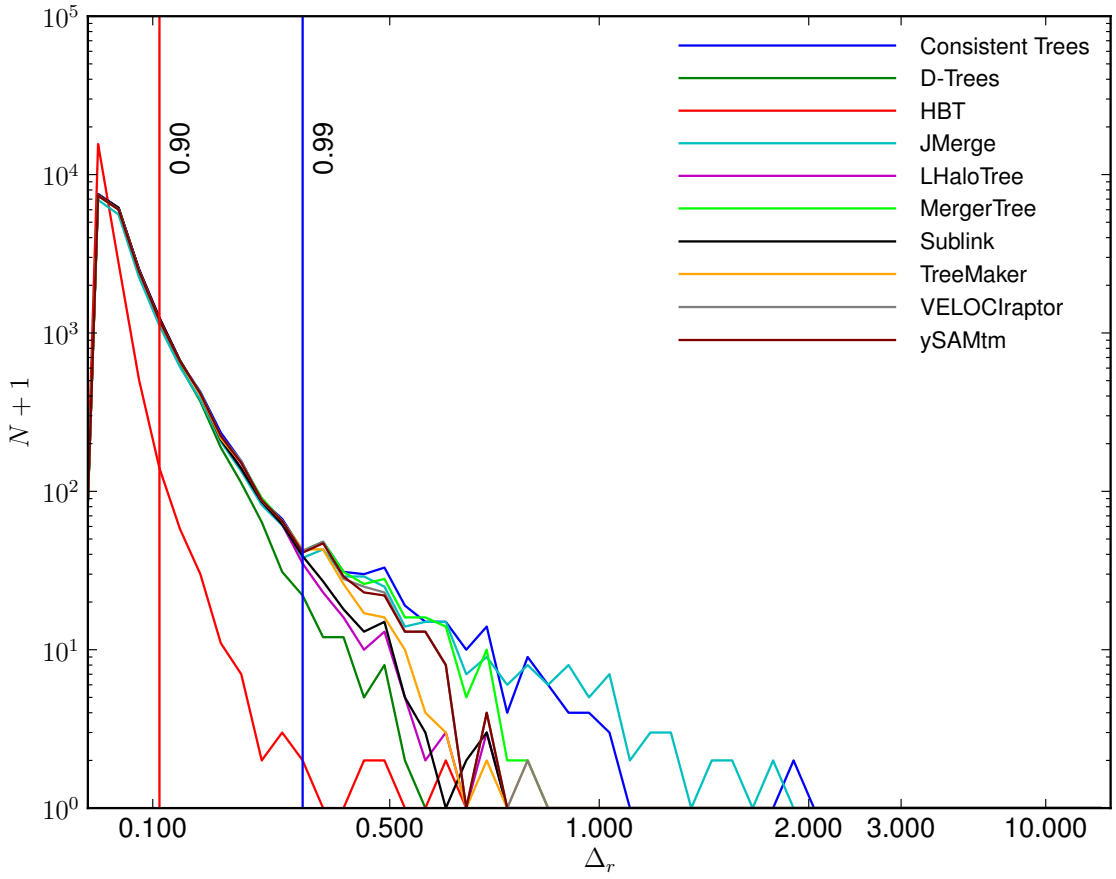

Figure 7.4: Histograms of the displacement statistic, Δ_r, for main haloes and their main progenitor for which both of them have $M_{200} > 10^{12}\,h^{-1}M_\odot$. The vertical lines show the 90$^{\text{th}}$ and 99$^{\text{th}}$ percentiles for MergerTree (but are approximately the same for all algorithms except HBT).

JMerge occasionally shows much larger deviations, suggesting that it does have a tendency to link together unassociated haloes. Consistent Trees also shows large outliers in this test and Figure 7.5 shows a typical example of how this comes about. Here we see an interaction in which the assignment of main halo alternates between successive snapshots:

- Most algorithms (top row) link together the visually correct group of particles and have small Δ_r, but will have a large fluctuation in halo mass along the main branch.
- JMerge requires smooth changes in mass and so it follows the main halo between Snapshots 58 & 59, leading to a large value of Δ_r.
- Consistent Trees follows the main branch across all three snapshots, giving large values of Δ_r for both links. It (correctly) fails to associate the top-right halo in

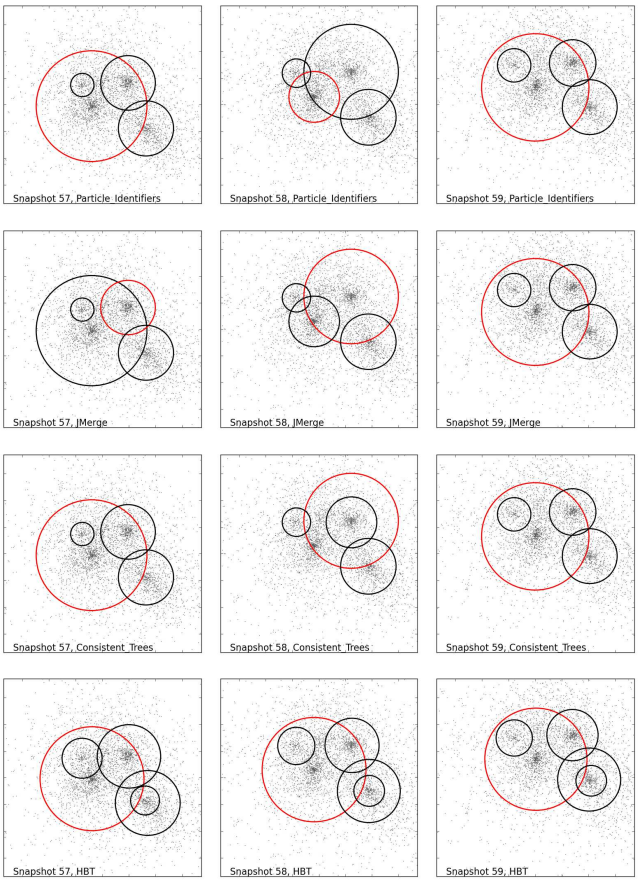

Figure 7.5: An example of a situation where the halo finder assigns main halos differently between snapshots. The red haloes in each row show the main branch of the largest halo on the right-hand side.

Snapshot 59 with the central one in Snapshot 58, so it removes the latter and creates a fake halo to take its place.

- HBT resolves the situation by creating a halo catalogue in which the mass evolution is smoother. It also inserts an extra subhalo on the bottom-right that is not returned by any of the other algorithms.

7.1.4 The loss of particles during halo growth

During mergers (and, indeed, during quiescent evolution) particles can be lost from haloes. As a measure of this, we use the statistic

$$\Delta_N = \frac{N_{\cup A_i} - N_{(\cup A_i)\cap B}}{N_{\cup A_i}}, \tag{7.2}$$

where, for a given halo B, the union runs over all direct progenitors, A_i. Here N is the number of particles in $\cup A_i$ and B or common to them both, as indicated by the subscript.

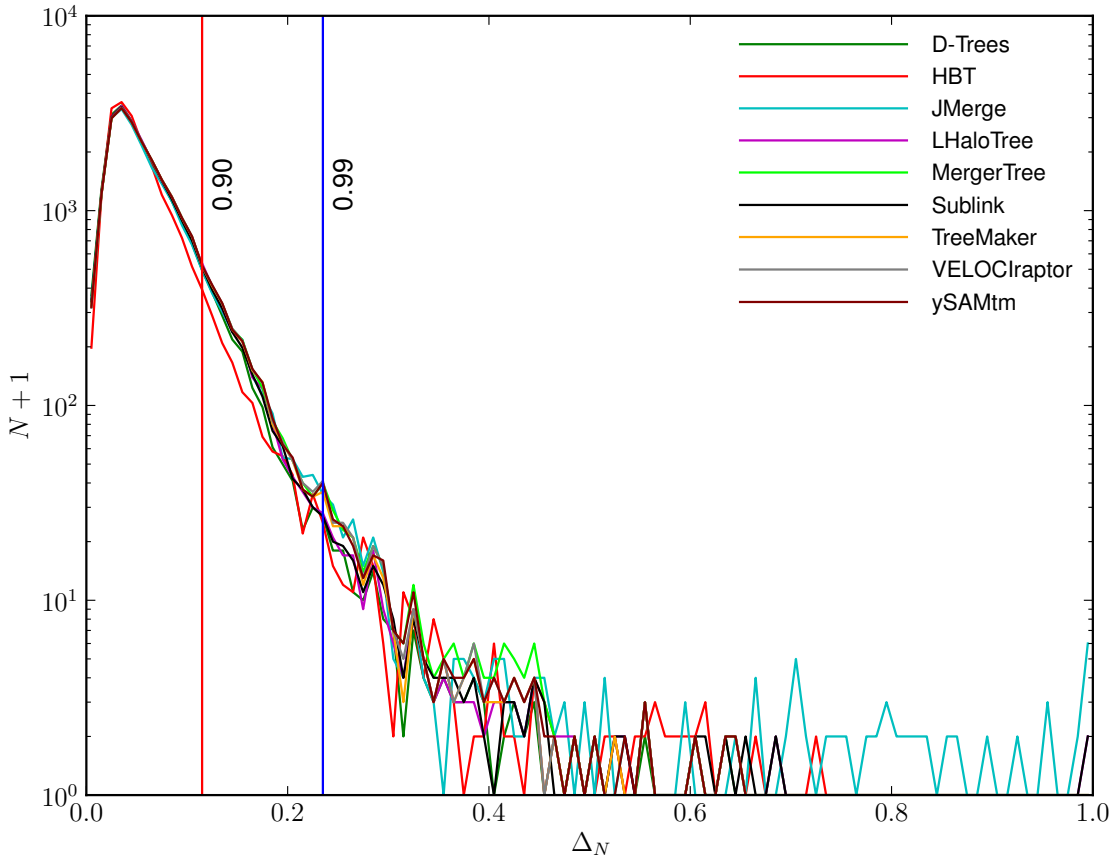

Figure 7.6: The distribution function of the fraction of lost particles, Δ_N for haloes along the main branch with $M_{200} > 10^{12}\,h^{-1}M_\odot$. The vertical lines show the 90$^{\rm th}$ and 99$^{\rm th}$ percentiles for MergerTree (but are approximately the same for all algorithms). Please note that Consistent Trees cannot be included in this test because the added halos specified by the code do not have particle information.

The distribution function of the fraction of lost particles, Δ_N for haloes along the main branch with $M_{200} > 10^{12}\,h^{-1}\mathrm{M}_\odot$ (corresponding to about 1000 particles) is shown in Figure 7.6. Note the extensive wing on this plot that extends to $\Delta_N = 0.4$. For small values of Δ_N, this is due to changes in the shape of the halo, and to natural particle orbits that results in material moving out across the radius (here R_{200}) used to define the edge of the halo. Large values of Δ_N can occur when haloes reduce their size significantly between snapshots. An example of this situation has already been shown in the third row of Figure 7.2 which illustrates how the halo finder alternates between allocating most of the mass to one or other of two haloes as they fly by one another.

All halo finders roughly agree on the number of haloes for which $\Delta_N < 0.4$, but there are signficant differences for larger values – these are most probably due to mis-identifications. It is perhaps not surprising that JMerge has occasional very poor matches,

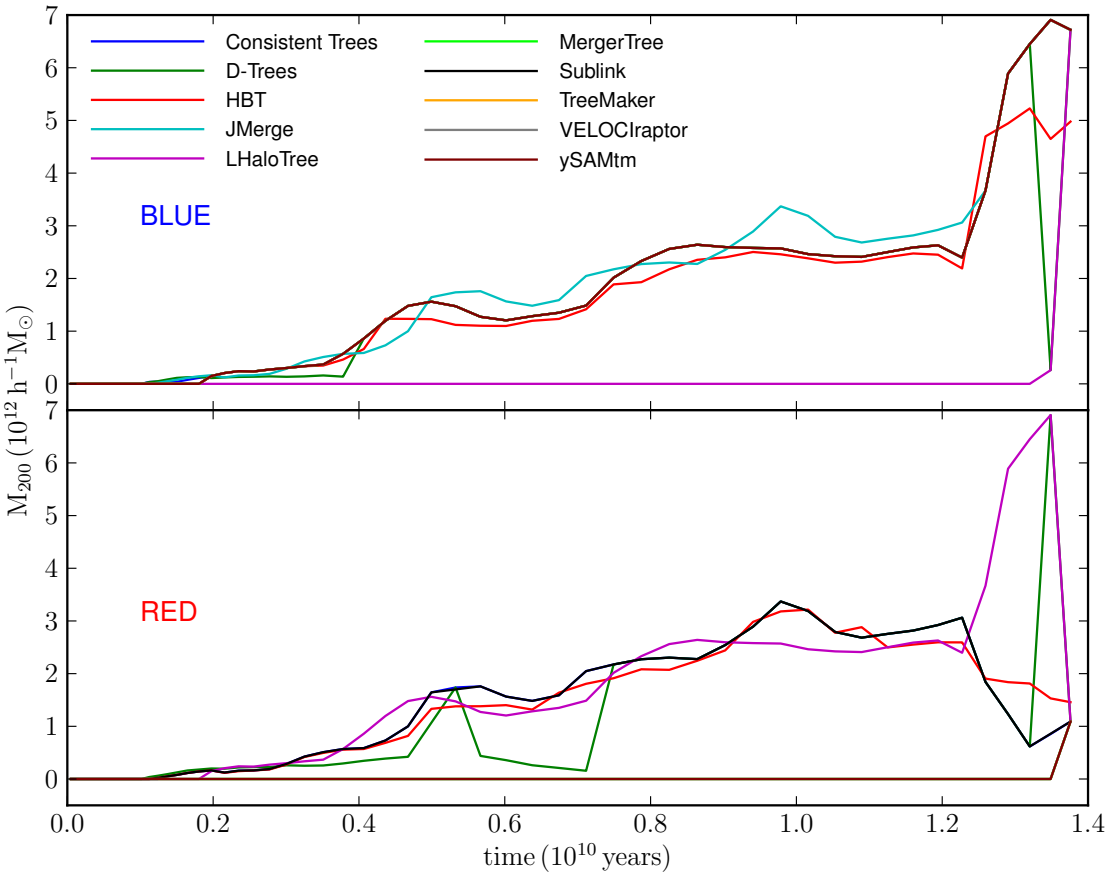

Figure 7.7: The mass history of the blue halo (top) and the red halo (bottom) in Figure 7.2 specified by each merger-tree code.

given that it does not use particle IDs, but rare examples of apparently erroneous links are found in many other algorithms too.

7.2 Mass growth

In this section we look at the mass evolution of haloes, primarily along their main branches, which is a key input for most SA models. While main haloes are expected to grow in mass through accretion and mergers, sub-haloes can lose mass through tidal stripping.

Consider first Figure 7.7 which shows the mass evolution along the main branch for the red and blue haloes illustrated Figure 7.2. The large mass fluctuations seen on the right-hand side of this plot correspond to the rightmost panels in Figure 7.2 and illustrate how poorly-constrained the mass evolution is during that merger – most SA models would struggle to cope with this kind of fluctuating mass behaviour.

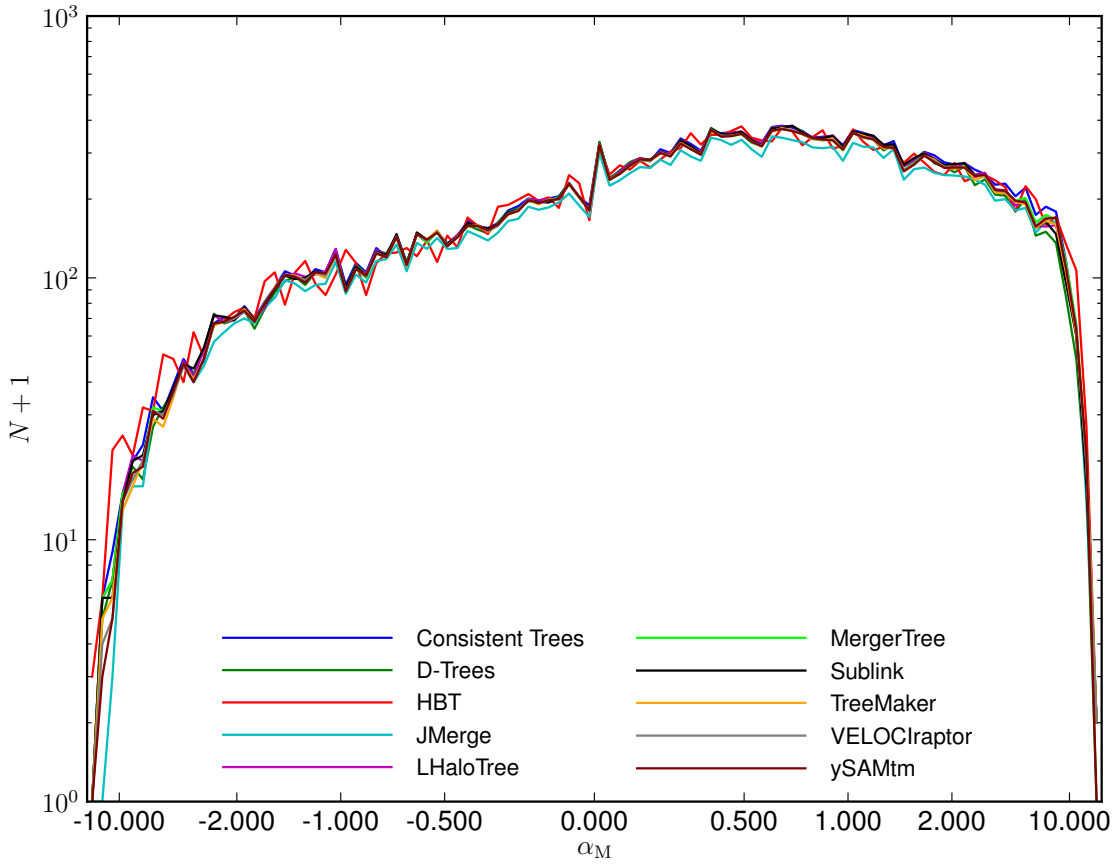

Figure 7.8: Distribution function of logarithmic mass growth, $\alpha_{\rm M}$ along halo main branches. We have included all pairs of haloes for which both the masses exceed 10^{12} $h^{-1}M_{\odot}$.

7.2.1 Mass growth along the halo main branch

The logarithmic growth rate of main branch haloes, $\mathrm{d}\log M/\mathrm{d}\log t$ is approximated discretely by

$$\frac{\mathrm{d}\log M}{\mathrm{d}\log t} \approx \alpha_{\rm M}(A,B) = \frac{(t_B + t_A)(M_B - M_A)}{(t_B - t_A)(M_B + M_A)}, \tag{7.3}$$

where M_A and M_B are the masses of a halo and its descendent at times t_A and t_B, respectively. The distribution function of $\alpha_{\rm M}$ is shown in Figure 7.8 for every pair of main-branch haloes for which the mass of each exceeds $10^{12}\,h^{-1}{\rm M}_{\odot}$ (corresponding to about 1000 particles).

As demonstrated in Figure 7.8, most of the time haloes are growing but there is a significant proportion of the time (about 30%) during which mass loss occurs. Such a large fraction is unlikely to be due to stripping (as this result is restricted to high-mass main-branch haloes) but some apparent mass loss can occur due to changes in the shape of haloes during their evolution, especially following a major merger.

Strong mass loss, however, is unphysical and is due to failures in the halo-finding and linking process, as illustrated in Figures 7.2, 7.5 & 7.7. The halo evolution seen in the rightmost columns of Figure 7.2 correspond to the wings in Figure 7.8.

7.2.2 Mass fluctuations of subhalo main branches

Abrupt fluctuations up and down in mass can be quantified with a statistic

$$\xi_M(k) = \arctan \alpha_M(k, k+1) - \arctan \alpha_M(k, k-1). \tag{7.4}$$

where α_M is as defined in Equation 7.3 and $k-1$, k & $k+1$ represent successive timesteps. This measures the change in the slope of the mass accretion rate between two consecutive steps and thus ranges from $-\pi$ to π. The main purpose of this statistic is to detect temporary mass fluctuations that occur either as a result of the natural growth process, or because of halo misidentification.

Large, negative values of ξ_M correspond to sharp temporary peaks in mass, and positive values to dips in mass. Somewhat surprisingly $|\xi_M|$ exceeds $\pi/3$ 10 per cent of the time, and $2\pi/3$ 1 per cent of the time. Thus strong mass variations are relatively common.

Note that the apparent discrepancy of HBT is because, for the purposes of this paper, they construct masses only from the supplied AHF halo catalogues. We have checked that, on applying HBT to the full simulation data, this discrepancy goes away.

7.3 Discussion

This paper summarises the results of a merger tree comparison project. The comparison was completed, and the paper drafted, in advance of the Sussing Merger Trees Workshop in Midhurst, Sussex in July 2013. The aim of the workshop was not only to compare the existing status of merger tree codes, but also to get people thinking about the desirable features of such codes, in particular for their use as backbones for SA modelling.

Ten different merger tree builders contributed to this comparison project, as listed in Table 6.1. Although many of these adopted similar approaches, no two gave identical results.

In order to enable the comparison, we desired that each merger tree code should use the same haloes as input. It soon became apparent that the halo finder can be intimately linked to the tree-builder itself, and so some tree-building codes needed modification to enable them to take part. For two of the codes (Consistent Trees & HBT), we had to allow modification of the halo catalogue. For this reason, and because the quality of a

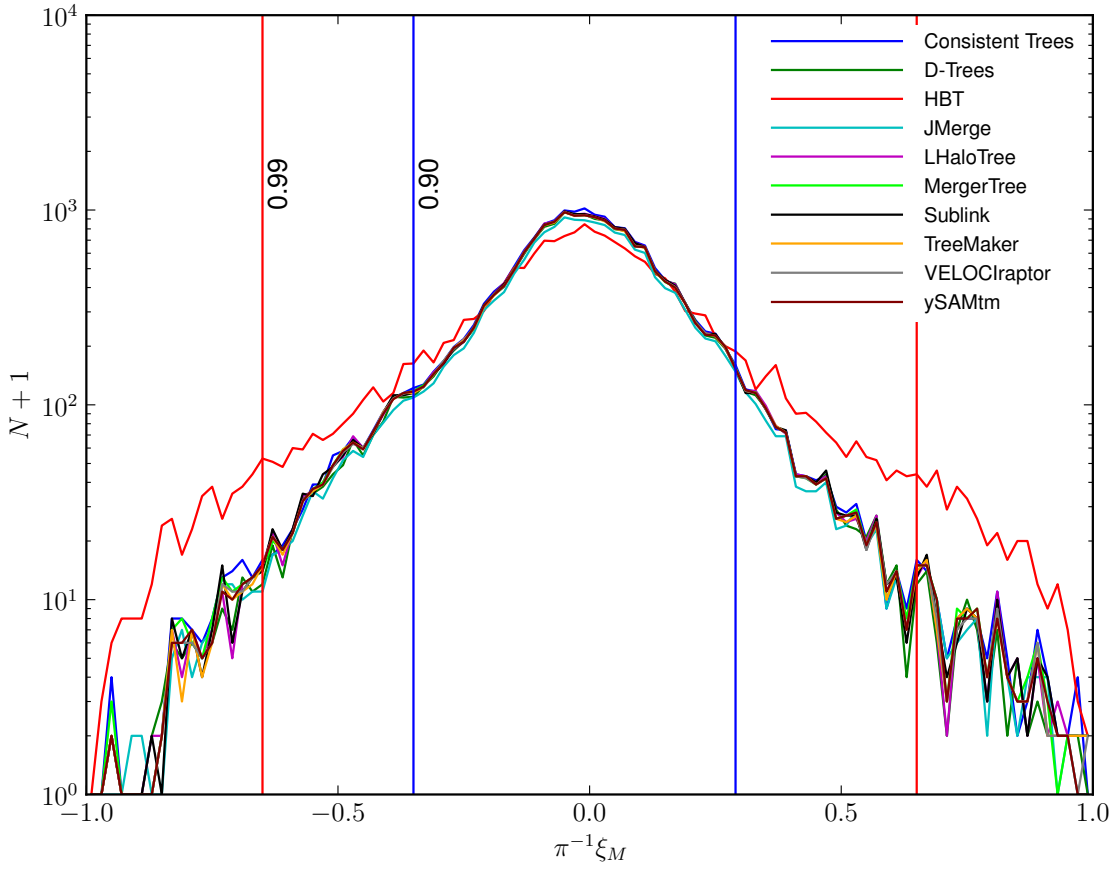

Figure 7.9: Mass fluctuations, ξ_M, for sets of 3 consecutive haloes along a main branch for which the mass of each exceeds $10^{12}\ h^{-1}M_\odot$. The vertical lines show the two-sided 90^{th} and 99^{th} percentiles for MergerTree (but are approximately the same for all algorithms except HBT). Note that the apparent discrepancy of HBT is because, for the purposes of this paper, they construct masses only from the supplied AHF halo catalogues. We have checked that, on applying HBT to the full simulation data, this discrepancy goes away.

merger tree depends in some unspecified way upon the particular scientific use to which it will be put, we avoid making conclusive statements here about which algorithms perform better than others.

In Section 6.2, we defined some terminology that we used throughout the paper. This proved essential to get everyone talking a common language (for example, some algorithms did not initially return merger *trees* at all, in the sense that every halo did not have a unique descendent). We encourage other members of the community to use the same nomenclature.

7.3.1 Summary of results

Here we present a brief summary of our findings:

- Imperfections in the halo finder can lead to great difficulties for tree-building algorithms. The particular halo finder that we used in this project was AHF, but we would expect similar behaviour with other halo finders and a study of this is presently under way.

- The temporary loss of a halo during the merger of two haloes (see, e.g. Figure 7.2) is disastrous for tree-building algorithms that examine only two adjacent snapshots. In such cases, it is possible for haloes containing over 1000 particles to apparently appear out of nothing between two adjacent snapshots.

- Although they were working with the same input halo catalogue, different algorithms varied in their ability to link together subhaloes, leading to significantly different branching ratios for the trees.

- Due to the limitations of the halo finder, codes that do not use particle IDs to link together haloes can occasionally produce clear mis-identifications (see, e.g. Figure 7.5).

- Even when haloes persist between snapshots, the halo finder will sometimes alter which of the two it treats as the main halo, and this can lead to large oscillations in mass. Different tree-builders handle this in different ways.

- The slope of the logarithmic mass growth curve, $\mathrm{dlog}\, M/\mathrm{dlog}\, t$ has a very broad distribution with a peak around 0.5 to 1 but extending beyond the range -10 to 10. Much of this is due to genuine fluctuations in mass, although the extremes are due to failures in the combined halo finder and tree builder.

We suggest that any optimal tree-building algorithm will require a high-quality input halo catalogue that minimises 'lost' haloes and mass fluctutations, and in addition will possess the following:

- the use of particle IDs to match haloes between snapshots;

- the ability to skip at least one, and preferably more, snapshots in order to recover subhaloes that are temporarily lost by the halo finder (for instance when they transit the centre of the host halo);

- the ability to cope with (and ideally smooth out) large, temporary fluctuations in halo mass.

7.4 Acknowledgements

The authors contributed in the following ways to this paper: CS, AK, FRP, AS, PAT organised this project. They designed the comparison, planned and organised the data, performed the analysis presented and wrote the paper. CS is a PhD student supervised by PAT. The other authors (as listed in Section 7.1) provided results and descriptions of their respective algorithms; they also helped to proof-read the paper.

Bibliography

Aghanim, N., Majumdar, S., and Silk, J. (2008). Secondary anisotropies of the CMB. *Reports on Progress in Physics*, 71(6):066902. 13

Alvarez, M. A., Bromm, V., and Shapiro, P. R. (2006). The H II Region of the First Star. ApJ, 639:621–632. 18

Alvarez, M. A., Busha, M., Abel, T., and Wechsler, R. H. (2009). Connecting Reionization to the Local Universe. ApJ, 703:L167–L171. 26

Aubert, D., Deparis, N., and Ocvirk, P. (2015). EMMA: an adaptive mesh refinement cosmological simulation code with radiative transfer. MNRAS, 454:1012–1037. 27

Aubert, D., Pichon, C., and Colombi, S. (2004). The origin and implications of dark matter anisotropic cosmic infall on L* haloes. 352:376–398. 78

Avila, S., Knebe, A., Pearce, F. R., Schneider, A., Srisawat, C., Thomas, P. A., Behroozi, P., Elahi, P. J., Han, J., Mao, Y.-Y., Onions, J., Rodriguez-Gomez, V., and Tweed, D. (2014). SUSSING MERGER TREES: the influence of the halo finder. MNRAS, 441:3488–3501. 31, 32, 68

Backer, D. C., Parsons, A., Bradley, R., Parashare, C., Gugliucci, N., Mastrantonio, E., Herne, D., Lynch, M., Wright, M., Werhimer, D., Carilli, C., Datta, A., and Aguirre, J. (2007). PAPER: The Precision Array To Probe The Epoch Of Reionization. In *American Astronomical Society Meeting Abstracts*, volume 39 of *Bulletin of the American Astronomical Society*, page 133.02. 14

Barkana, R. and Loeb, A. (1999). The Photoevaporation of Dwarf Galaxies during Reionization. ApJ, 523:54–65. 27

Barkana, R. and Loeb, A. (2001). In the beginning: the first sources of light and the reionization of the universe. Phys. Rep., 349:125–238. 11, 12, 16, 18

Barnes, J. and Hut, P. (1986). A hierarchical O(N log N) force-calculation algorithm. Nature, 324:446–449. 4

Baugh, C. M., Cole, S., and Frenk, C. S. (1996). Evolution of the Hubble sequence in hierarchical models for galaxy formation. MNRAS, 283:1361–1378. 5

Becker, G. D., Bolton, J. S., and Lidz, A. (2015a). Reionization and high-redshift galaxies: the view from quasar absorption lines. *ArXiv e-prints*. 10, 13

Becker, G. D., Bolton, J. S., Madau, P., Pettini, M., Ryan-Weber, E. V., and Venemans, B. P. (2015b). Evidence of patchy hydrogen reionization from an extreme Lyα trough below redshift six. MNRAS, 447:3402–3419. 25

Becker, R. H., Fan, X., White, R. L., Strauss, M. A., Narayanan, V. K., Lupton, R. H., Gunn, J. E., Annis, J., Bahcall, N. A., Brinkmann, J., Connolly, A. J., Csabai, I., Czarapata, P. C., Doi, M., Heckman, T. M., Hennessy, G. S., Ivezić, Ž., Knapp, G. R., Lamb, D. Q., McKay, T. A., Munn, J. A., Nash, T., Nichol, R., Pier, J. R., Richards, G. T., Schneider, D. P., Stoughton, C., Szalay, A. S., Thakar, A. R., and York, D. G. (2001). Evidence for Reionization at z˜6: Detection of a Gunn-Peterson Trough in a z=6.28 Quasar. AJ, 122:2850–2857. 12, 25

Behroozi, P. S., Wechsler, R. H., and Wu, H.-Y. (2013a). The ROCKSTAR Phase-space Temporal Halo Finder and the Velocity Offsets of Cluster Cores. ApJ, 762:109. 6

Behroozi, P. S., Wechsler, R. H., Wu, H.-Y., Busha, M. T., Klypin, A. A., and Primack, J. R. (2013b). Gravitationally Consistent Halo Catalogs and Merger Trees for Precision Cosmology. 763:18. 73, 74

Benson, A. J., Frenk, C. S., Baugh, C. M., Cole, S., and Lacey, C. G. (2003). The effects of photoionization on galaxy formation - III. Environmental dependence in the luminosity function. MNRAS, 343:679–691. 127

Benson, A. J., Sugiyama, N., Nusser, A., and Lacey, C. G. (2006). The epoch of reionization. MNRAS, 369:1055–1080. 27

Bergvall, N., Zackrisson, E., Andersson, B.-G., Arnberg, D., Masegosa, J., and Östlin, G. (2006). First detection of Lyman continuum escape from a local starburst galaxy. I. Observations of the luminous blue compact galaxy Haro 11 with the Far Ultraviolet Spectroscopic Explorer (FUSE). A&A, 448:513–524. 17

Binney, J. and Tremaine, S. (1987). *Galactic dynamics*. 130

Blumenthal, G. R., Faber, S. M., Primack, J. R., and Rees, M. J. (1984). Formation of galaxies and large-scale structure with cold dark matter. Nature, 311:517–525. 3

Bolton, J. S. and Haehnelt, M. G. (2007). The observed ionization rate of the intergalactic medium and the ionizing emissivity at z ¿= 5: evidence for a photon-starved and extended epoch of reionization. MNRAS, 382:325–341. 16, 25, 42

Bond, J. R., Cole, S., Efstathiou, G., and Kaiser, N. (1991). Excursion set mass functions for hierarchical Gaussian fluctuations. ApJ, 379:440–460. 5, 9

Bond, J. R. and Efstathiou, G. (1984). Cosmic background radiation anisotropies in universes dominated by nonbaryonic dark matter. ApJ, 285:L45–L48. 13

Bouwens, R. J., Illingworth, G. D., Franx, M., and Ford, H. (2007). UV Luminosity Functions at z˜4, 5, and 6 from the Hubble Ultra Deep Field and Other Deep Hubble Space Telescope ACS Fields: Evolution and Star Formation History. ApJ, 670:928–958. 38

Bouwens, R. J., Illingworth, G. D., González, V., Labbé, I., Franx, M., Conselice, C. J., Blakeslee, J., van Dokkum, P., Holden, B., Magee, D., Marchesini, D., and Zheng, W. (2010a). z ˜ 7 Galaxy Candidates from NICMOS Observations Over the HDF-South and the CDF-South and HDF-North Goods Fields. ApJ, 725:1587–1599. 38

Bouwens, R. J., Illingworth, G. D., Oesch, P. A., Caruana, J., Holwerda, B., Smit, R., and Wilkins, S. (2015a). Cosmic Reionization after Planck: The Derived Growth of the Ionizing Background now matches the Growth of the Galaxy UV Luminosity Density. *ArXiv e-prints*. 17, 38

Bouwens, R. J., Illingworth, G. D., Oesch, P. A., Labbé, I., Trenti, M., van Dokkum, P., Franx, M., Stiavelli, M., Carollo, C. M., Magee, D., and Gonzalez, V. (2011). Ultraviolet Luminosity Functions from 132 z ˜ 7 and z ˜ 8 Lyman-break Galaxies in the Ultra-deep HUDF09 and Wide-area Early Release Science WFC3/IR Observations. ApJ, 737:90. 26

Bouwens, R. J., Illingworth, G. D., Oesch, P. A., Stiavelli, M., van Dokkum, P., Trenti, M., Magee, D., Labbé, I., Franx, M., Carollo, C. M., and Gonzalez, V. (2010b). Discovery of z ˜ 8 Galaxies in the Hubble Ultra Deep Field from Ultra-Deep WFC3/IR Observations. ApJ, 709:L133–L137. 26

Bouwens, R. J., Illingworth, G. D., Oesch, P. A., Trenti, M., Labbé, I., Bradley, L., Carollo, M., van Dokkum, P. G., Gonzalez, V., Holwerda, B., Franx, M., Spitler, L., Smit, R., and Magee, D. (2015b). UV Luminosity Functions at Redshifts z 4 to z 10: 10,000 Galaxies from HST Legacy Fields. ApJ, 803:34. 26, 38, 63

Bouwens, R. J., Illingworth, G. D., Oesch, P. A., Trenti, M., Labbé, I., Franx, M., Stiavelli, M., Carollo, C. M., van Dokkum, P., and Magee, D. (2012). Lower-luminosity Galaxies Could Reionize the Universe: Very Steep Faint-end Slopes to the UV Luminosity Functions at z ¿= 5-8 from the HUDF09 WFC3/IR Observations. ApJ, 752:L5. 17

Bower, R. G., Benson, A. J., Malbon, R., Helly, J. C., Frenk, C. S., Baugh, C. M., Cole, S., and Lacey, C. G. (2006). Breaking the hierarchy of galaxy formation. MNRAS, 370:645–655. 27

Bowler, R. A. A., Dunlop, J. S., McLure, R. J., Rogers, A. B., McCracken, H. J., Milvang-Jensen, B., Furusawa, H., Fynbo, J. P. U., Taniguchi, Y., Afonso, J., Bremer, M. N., and Le Fèvre, O. (2014). The bright end of the galaxy luminosity function at z∼7: before the onset of mass quenching? MNRAS, 440:2810–2842. 38

Boylan-Kolchin, M., Springel, V., White, S. D. M., Jenkins, A., and Lemson, G. (2009). Resolving cosmic structure formation with the Millennium-II Simulation. MNRAS, 398:1150–1164. 10

Brentjens, M., Koopmans, L. V. E., de Bruyn, A. G., and Zaroubi, S. (2011). The Low-Frequency Array (LOFAR) and EoR Key-Science Project. In *American Astronomical Society Meeting Abstracts 217*, volume 43 of *Bulletin of the American Astronomical Society*, page 107.04. 14

Bromm, V. and Larson, R. B. (2004). The First Stars. ARA&A, 42:79–118. 11, 12

Bromm, V. and Loeb, A. (2006). High-Redshift Gamma-Ray Bursts from Population III Progenitors. ApJ, 642:382–388. 16

Bruzual, G. and Charlot, S. (2003). Stellar population synthesis at the resolution of 2003. MNRAS, 344:1000–1028. 133

Bunker, A. J., Wilkins, S., Ellis, R. S., Stark, D. P., Lorenzoni, S., Chiu, K., Lacy, M., Jarvis, M. J., and Hickey, S. (2010). The contribution of high-redshift galaxies to cosmic reionization: new results from deep WFC3 imaging of the Hubble Ultra Deep Field. MNRAS, 409:855–866. 26

Cattaneo, A., Blaizot, J., Weinberg, D. H., Kereš, D., Colombi, S., Davé, R., Devriendt, J., Guiderdoni, B., and Katz, N. (2007). Accretion, feedback and galaxy bimodality: a comparison of the GalICS semi-analytic model and cosmological SPH simulations. MNRAS, 377:63–76. 9

Chabrier, G. (2003). Galactic Stellar and Substellar Initial Mass Function. PASP, 115:763–795. 33

Chen, X., Cooray, A., Yoshida, N., and Sugiyama, N. (2003). Can non-Gaussian cosmological models explain the WMAP high optical depth for reionization? MNRAS, 346:L31–L35. 18

Choudhury, T. R. and Ferrara, A. (2006). Physics of Cosmic Reionization. *ArXiv Astrophysics e-prints*. 11, 12

Choudhury, T. R., Haehnelt, M. G., and Regan, J. (2009). Inside-out or outside-in: the topology of reionization in the photon-starved regime suggested by Lyα forest data. MNRAS, 394:960–977. 26, 34, 35, 52

Ciardi, B. and Ferrara, A. (2005). The First Cosmic Structures and Their Effects. Space Sci. Rev., 116:625–705. 11, 12

Clay, S. J., Thomas, P. A., Wilkins, S. M., and Henriques, B. M. B. (2015). Galaxy formation in the Planck cosmology - III. The high-redshift universe. MNRAS, 451:2692–2702. 27, 31

Cole, S. (1991). Modeling galaxy formation in evolving dark matter halos. ApJ, 367:45–53. 5, 9

Cole, S., Aragon-Salamanca, A., Frenk, C. S., Navarro, J. F., and Zepf, S. E. (1994). A Recipe for Galaxy Formation. MNRAS, 271:781. 9

Cole, S., Lacey, C. G., Baugh, C. M., and Frenk, C. S. (2000). Hierarchical galaxy formation. MNRAS, 319:168–204. 9

Cordes, J. M. (2005). The Square Kilometer Array: Key Science and Technology Development. In Kassim, N., Perez, M., Junor, W., and Henning, P., editors, *From Clark Lake to the Long Wavelength Array: Bill Erickson's Radio Science*, volume 345 of *Astronomical Society of the Pacific Conference Series*, page 461. 15

Cortese, L., Davies, J. I., Pohlen, M., Baes, M., Bendo, G. J., Bianchi, S., Boselli, A., De Looze, I., Fritz, J., Verstappen, J., Bomans, D. J., Clemens, M., Corbelli, E., Dariush, A., di Serego Alighieri, S., Fadda, D., Garcia-Appadoo, D. A., Gavazzi, G., Giovanardi, C., Grossi, M., Hughes, T. M., Hunt, L. K., Jones, A. P., Madden, S., Pierini, D., Sabatini, S., Smith, M. W. L., Vlahakis, C., Xilouris, E. M., and Zibetti, S. (2010). The Herschel Virgo Cluster Survey . II. Truncated dust disks in H I-deficient spirals. A&A, 518:L49. 14

Crill, B. P., Ade, P. A. R., Artusa, D. R., Bhatia, R. S., Bock, J. J., Boscaleri, A., Cardoni, P., Church, S. E., Coble, K., de Bernardis, P., de Troia, G., Farese, P., Ganga, K. M., Giacometti, M., Haynes, C. V., Hivon, E., Hristov, V. V., Iacoangeli, A., Jones, W. C., Lange, A. E., Martinis, L., Masi, S., Mason, P. V., Mauskopf, P. D., Miglio, L., Montroy, T., Netterfield, C. B., Paine, C. G., Pascale, E., Piacentini, F., Polenta, G., Pongetti, F., Romeo, G., Ruhl, J. E., Scaramuzzi, F., Sforna, D., and Turner, A. D. (2003). BOOMERANG: A Balloon-borne Millimeter-Wave Telescope and Total Power Receiver for Mapping Anisotropy in the Cosmic Microwave Background. ApJS, 148:527–541. 2

Crocce, M., Pueblas, S., and Scoccimarro, R. (2006). Transients from initial conditions in cosmological simulations. MNRAS, 373:369–381. 4, 10, 29

Croton, D. J., Springel, V., White, S. D. M., De Lucia, G., Frenk, C. S., Gao, L., Jenkins, A., Kauffmann, G., Navarro, J. F., and Yoshida, N. (2006). The many lives of active galactic nuclei: cooling flows, black holes and the luminosities and colours of galaxies. MNRAS, 365:11–28. 9, 27, 31, 127

Davis, M., Efstathiou, G., Frenk, C. S., and White, S. D. M. (1985a). The evolution of large-scale structure in a universe dominated by cold dark matter. ApJ, 292:371–394. 5, 30

Davis, M., Efstathiou, G., Frenk, C. S., and White, S. D. M. (1985b). The evolution of large-scale structure in a universe dominated by cold dark matter. 292:371–394. 78

De Lucia, G. and Blaizot, J. (2007). The hierarchical formation of the brightest cluster galaxies. MNRAS, 375:2–14. 10, 27, 31, 130

De Lucia, G., Kauffmann, G., Springel, V., White, S. D. M., Lanzoni, B., Stoehr, F., Tormen, G., and Yoshida, N. (2004a). Substructures in cold dark matter haloes. 348:333–344. 125, 129

De Lucia, G., Kauffmann, G., and White, S. D. M. (2004b). Chemical enrichment of the intracluster and intergalactic medium in a hierarchical galaxy formation model. MNRAS, 349:1101–1116. 31

De Lucia, G., Springel, V., White, S. D. M., Croton, D., and Kauffmann, G. (2006). The formation history of elliptical galaxies. MNRAS, 366:499–509. 9

Deharveng, J.-M., Buat, V., Le Brun, V., Milliard, B., Kunth, D., Shull, J. M., and Gry, C. (2001). Constraints on the Lyman continuum radiation from galaxies: First results with FUSE on Mrk 54. A&A, 375:805–813. 17

Dekel, A. and Silk, J. (1986). The origin of dwarf galaxies, cold dark matter, and biased galaxy formation. ApJ, 303:39–55. 127

Dickey, J. M. and Lockman, F. J. (1990). H I in the Galaxy. ARA&A, 28:215–261. 14

Dijkstra, M., Haiman, Z., and Loeb, A. (2004a). A Limit from the X-Ray Background on the Contribution of Quasars to Reionization. ApJ, 613:646–654. 16

Dijkstra, M., Haiman, Z., Rees, M. J., and Weinberg, D. H. (2004b). Photoionization Feedback in Low-Mass Galaxies at High Redshift. ApJ, 601:666–675. 31, 124

Dillon, J. S., Liu, A., Williams, C. L., Hewitt, J. N., Tegmark, M., Morgan, E. H., Levine, A. M., Morales, M. F., Tingay, S. J., Bernardi, G., Bowman, J. D., Briggs, F. H., Cappallo, R. C., Emrich, D., Mitchell, D. A., Oberoi, D., Prabu, T., Wayth, R., and Webster, R. L. (2014). Overcoming real-world obstacles in 21 cm power spectrum estimation: A method demonstration and results from early Murchison Widefield Array data. Phys. Rev. D, 89(2):023002. 15

Dixon, K. L., Iliev, I. T., Mellema, G., Ahn, K., and Shapiro, P. R. (2016). The large-scale observational signatures of low-mass galaxies during reionization. MNRAS, 456:3011–3029. 29, 36, 52, 64, 65

Djorgovski, S. G., Castro, S., Stern, D., and Mahabal, A. A. (2001). On the Threshold of the Reionization Epoch. ApJ, 560:L5–L8. 12

Duffy, A. R., Wyithe, J. S. B., Mutch, S. J., and Poole, G. B. (2014). Low-mass galaxy formation and the ionizing photon budget during reionization. MNRAS, 443:3435–3443. 26

Duncan, K., Conselice, C. J., Mortlock, A., Hartley, W. G., Guo, Y., Ferguson, H. C., Davé, R., Lu, Y., Ownsworth, J., Ashby, M. L. N., Dekel, A., Dickinson, M., Faber, S., Giavalisco, M., Grogin, N., Kocevski, D., Koekemoer, A., Somerville, R. S., and White, C. E. (2014). The mass evolution of the first galaxies: stellar mass functions and star formation rates at $4 < z < 7$ in the CANDELS GOODS-South field. MNRAS, 444:2960–2984. 11, 26, 38, 47, 48, 49, 63

Efstathiou, G. (1992). Suppressing the formation of dwarf galaxies via photoionization. MNRAS, 256:43P–47P. 31, 45

Efstathiou, G., Davis, M., White, S. D. M., and Frenk, C. S. (1985). Numerical techniques for large cosmological N-body simulations. ApJS, 57:241–260. 4

Fan, X., Narayanan, V. K., Strauss, M. A., White, R. L., Becker, R. H., Pentericci, L., and Rix, H.-W. (2002). Evolution of the Ionizing Background and the Epoch of Reionization from the Spectra of z˜6 Quasars. AJ, 123:1247–1257. 12, 13

Fan, X., Strauss, M. A., Becker, R. H., White, R. L., Gunn, J. E., Knapp, G. R., Richards, G. T., Schneider, D. P., Brinkmann, J., and Fukugita, M. (2006). Constraining the Evolution of the Ionizing Background and the Epoch of Reionization with z˜6 Quasars. II. A Sample of 19 Quasars. AJ, 132:117–136. 12, 25

Fernández-Soto, A., Lanzetta, K. M., and Chen, H.-W. (2003). Mission: impossible (escape from the Lyman limit). MNRAS, 342:1215–1221. 17

Ferrara, A. and Pandolfi, S. (2014). Reionization of the Intergalactic Medium. *ArXiv e-prints*. 17, 18

Field, G. B. (1958). Excitation of the Hydrogen 21-CM Line. *Proceedings of the IRE*, 46:240–250. 14

Field, G. B. (1959). The Time Relaxation of a Resonance-Line Profile. ApJ, 129:551. 14

Finkelstein, S. L., Papovich, C., Giavalisco, M., Reddy, N. A., Ferguson, H. C., Koekemoer, A. M., and Dickinson, M. (2010). On the Stellar Populations and Evolution of Star-forming Galaxies at $6.3 < z <= 8.6$. ApJ, 719:1250–1273. 26

Finkelstein, S. L., Papovich, C., Ryan, R. E., Pawlik, A. H., Dickinson, M., Ferguson, H. C., Finlator, K., Koekemoer, A. M., Giavalisco, M., Cooray, A., Dunlop, J. S., Faber, S. M., Grogin, N. A., Kocevski, D. D., and Newman, J. A. (2012a). CANDELS:

The Contribution of the Observed Galaxy Population to Cosmic Reionization. ApJ, 758:93. 26, 42

Finkelstein, S. L., Papovich, C., Salmon, B., Finlator, K., Dickinson, M., Ferguson, H. C., Giavalisco, M., Koekemoer, A. M., Reddy, N. A., Bassett, R., Conselice, C. J., Dunlop, J. S., Faber, S. M., Grogin, N. A., Hathi, N. P., Kocevski, D. D., Lai, K., Lee, K.-S., McLure, R. J., Mobasher, B., and Newman, J. A. (2012b). Candels: The Evolution of Galaxy Rest-frame Ultraviolet Colors from z = 8 to 4. ApJ, 756:164. 26

Finkelstein, S. L., Ryan, Jr., R. E., Papovich, C., Dickinson, M., Song, M., Somerville, R. S., Ferguson, H. C., Salmon, B., Giavalisco, M., Koekemoer, A. M., Ashby, M. L. N., Behroozi, P., Castellano, M., Dunlop, J. S., Faber, S. M., Fazio, G. G., Fontana, A., Grogin, N. A., Hathi, N., Jaacks, J., Kocevski, D. D., Livermore, R., McLure, R. J., Merlin, E., Mobasher, B., Newman, J. A., Rafelski, M., Tilvi, V., and Willner, S. P. (2015). The Evolution of the Galaxy Rest-frame Ultraviolet Luminosity Function over the First Two Billion Years. ApJ, 810:71. 26

Finlator, K., Davé, R., and Özel, F. (2011). Galactic Outflows and Photoionization Heating in the Reionization Epoch. ApJ, 743:169. 26

Fontanot, F. and Monaco, P. (2010). The active and passive populations of extremely red objects. MNRAS, 405:705–717. 133

Fosbury, R. A. E., Villar-Martín, M., Humphrey, A., Lombardi, M., Rosati, P., Stern, D., Hook, R. N., Holden, B. P., Stanford, S. A., Squires, G. K., Rauch, M., and Sargent, W. L. W. (2003). Massive Star Formation in a Gravitationally Lensed H II Galaxy at z = 3.357. ApJ, 596:797–809. 18

Furlanetto, S. R. (2015). The 21-cm line as a probe of reionization. 14

Furlanetto, S. R. and Loeb, A. (2003). Metal Absorption Lines as Probes of the Intergalactic Medium Prior to the Reionization Epoch. ApJ, 588:18–34. 25

Furlanetto, S. R., Oh, S. P., and Briggs, F. H. (2006). Cosmology at low frequencies: The 21 cm transition and the high-redshift Universe. Phys. Rep., 433:181–301. 11, 12, 14

Furlanetto, S. R., Zaldarriaga, M., and Hernquist, L. (2004). The Growth of H II Regions During Reionization. ApJ, 613:1–15. 34

Furlong, M., Bower, R. G., Theuns, T., Schaye, J., Crain, R. A., Schaller, M., Dalla Vecchia, C., Frenk, C. S., McCarthy, I. G., Helly, J., Jenkins, A., and Rosas-Guevara,

Y. M. (2015). Evolution of galaxy stellar masses and star formation rates in the EAGLE simulations. MNRAS, 450:4486–4504. 26

Geil, P. M. and Wyithe, J. S. B. (2008). The impact of a percolating IGM on redshifted 21-cm observations of quasar HII regions. MNRAS, 386:1683–1694. 26

Genel, S., Vogelsberger, M., Springel, V., Sijacki, D., Nelson, D., Snyder, G., Rodriguez-Gomez, V., Torrey, P., and Hernquist, L. (2014). Introducing the Illustris project: the evolution of galaxy populations across cosmic time. MNRAS, 445:175–200. 26

Gill, S. P., Knebe, A., and Gibson, B. K. (2004). The evolution of substructureI. a new identification method. 351(2):399–409. 70

Giroux, M. L. and Shapiro, P. R. (1996). The Reionization of the Intergalactic Medium and Its Observational Consequences. ApJS, 102:191. 16

Gnedin, N. Y. (2000). Effect of Reionization on Structure Formation in the Universe. ApJ, 542:535–541. 18, 21, 27, 32, 36

Gnedin, N. Y. (2014). Cosmic Reionization on Computers. I. Design and Calibration of Simulations. ApJ, 793:29. 27

Gnedin, N. Y. and Abel, T. (2001). Multi-dimensional cosmological radiative transfer with a Variable Eddington Tensor formalism. NewA, 6:437–455. 25

Gnedin, N. Y. and Hui, L. (1998). Probing the Universe with the Lyalpha forest - I. Hydrodynamics of the low-density intergalactic medium. MNRAS, 296:44–55. 21

González, V., Labbé, I., Bouwens, R. J., Illingworth, G., Franx, M., and Kriek, M. (2011). Evolution of Galaxy Stellar Mass Functions, Mass Densities, and Mass-to-light Ratios from z ˜ 7 to z ˜ 4. ApJ, 735:L34. 11, 47, 49

Gonzalez-Perez, V., Lacey, C. G., Baugh, C. M., Lagos, C. D. P., Helly, J., Campbell, D. J. R., and Mitchell, P. D. (2014). How sensitive are predicted galaxy luminosities to the choice of stellar population synthesis model? MNRAS, 439:264–283. 133

Graziani, L., Salvadori, S., Schneider, R., Kawata, D., de Bennassuti, M., and Maselli, A. (2015). Galaxy formation with radiative and chemical feedback. MNRAS, 449:3137–3148. 27

Guiderdoni, B. and Rocca-Volmerange, B. (1987). A model of spectrophotometric evolution for high-redshift galaxies. A&A, 186:1–21. 134

Gunn, J. E. and Peterson, B. A. (1965). On the Density of Neutral Hydrogen in Intergalactic Space. ApJ, 142:1633–1641. 12

Guo, Q., White, S., Angulo, R. E., Henriques, B., Lemson, G., Boylan-Kolchin, M., Thomas, P., and Short, C. (2013). Galaxy formation in WMAP1 and WMAP7 cosmologies. MNRAS, 428:1351–1365. 10, 27, 31

Guo, Q., White, S., Boylan-Kolchin, M., De Lucia, G., Kauffmann, G., Lemson, G., Li, C., Springel, V., and Weinmann, S. (2011). From dwarf spheroidals to cD galaxies: simulating the galaxy population in a ΛCDM cosmology. MNRAS, 413:101–131. 10, 27, 31, 125, 127, 128, 129, 130, 131, 132

Haardt, F. and Madau, P. (1996). Radiative Transfer in a Clumpy Universe. II. The Ultraviolet Extragalactic Background. ApJ, 461:20. 27

Haardt, F. and Madau, P. (2012). Radiative Transfer in a Clumpy Universe. IV. New Synthesis Models of the Cosmic UV/X-Ray Background. ApJ, 746:125. 27

Haiman, Z. and Holder, G. P. (2003). The Reionization History at High Redshifts. I. Physical Models and New Constraints from Cosmic Microwave Background Polarization. ApJ, 595:1–12. 34

Han, J., Jing, Y. P., Wang, H., and Wang, W. (2012a). Resolving subhaloes' lives with the Hierarchical Bound-Tracing algorithm. MNRAS, 427:2437–2449. 6

Han, J., Jing, Y. P., Wang, H., and Wang, W. (2012b). Resolving subhaloes' lives with the Hierarchical Bound-Tracing algorithm. 427:2437–2449. 75

Hansen, S. H. and Haiman, Z. (2004). Do We Need Stars to Reionize the Universe at High Redshifts? Early Reionization by Decaying Heavy Sterile Neutrinos. ApJ, 600:26–31. 18

Harnois-Déraps, J., Pen, U.-L., Iliev, I. T., Merz, H., Emberson, J. D., and Desjacques, V. (2013). High-performance P^3M N-body code: CUBEP^{3}M. MNRAS, 436:540–559. 29

Hatton, S., Devriendt, J. E. G., Ninin, S., Bouchet, F. R., Guiderdoni, B., and Vibert, D. (2003). GALICS- I. A hybrid N-body/semi-analytic model of hierarchical galaxy formation. 343:75–106. 78

Heckman, T. M., Sembach, K. R., Meurer, G. R., Leitherer, C., Calzetti, D., and Martin, C. L. (2001). On the Escape of Ionizing Radiation from Starbursts. ApJ, 558:56–62. 17

Henriques, B., Maraston, C., Monaco, P., Fontanot, F., Menci, N., De Lucia, G., and Tonini, C. (2011). The effect of thermally pulsating asymptotic giant branch stars on the evolution of the rest-frame near-infrared galaxy luminosity function. MNRAS, 415:3571–3579. 133

Henriques, B. M. B., Thomas, P. A., Oliver, S., and Roseboom, I. (2009). Monte Carlo Markov Chain parameter estimation in semi-analytic models of galaxy formation. MNRAS, 396:535–547. 10, 38

Henriques, B. M. B., White, S. D. M., Lemson, G., Thomas, P. A., Guo, Q., Marleau, G.-D., and Overzier, R. A. (2012). Confronting theoretical models with the observed evolution of the galaxy population out to z= 4. MNRAS, 421:2904–2916. 133

Henriques, B. M. B., White, S. D. M., Thomas, P. A., Angulo, R., Guo, Q., Lemson, G., Springel, V., and Overzier, R. (2015). Galaxy formation in the Planck cosmology - I. Matching the observed evolution of star formation rates, colours and stellar masses. MNRAS, 451:2663–2680. 9, 10, 27, 31, 40, 63, 124, 127, 128, 129, 131, 133

Henriques, B. M. B., White, S. D. M., Thomas, P. A., Angulo, R. E., Guo, Q., Lemson, G., and Springel, V. (2013). Simulations of the galaxy population constrained by observations from z = 3 to the present day: implications for galactic winds and the fate of their ejecta. MNRAS, 431:3373–3395. 9, 31, 40, 128

Hinshaw, G., Larson, D., Komatsu, E., Spergel, D. N., Bennett, C. L., Dunkley, J., Nolta, M. R., Halpern, M., Hill, R. S., Odegard, N., Page, L., Smith, K. M., Weiland, J. L., Gold, B., Jarosik, N., Kogut, A., Limon, M., Meyer, S. S., Tucker, G. S., Wollack, E., and Wright, E. L. (2013). Nine-year Wilkinson Microwave Anisotropy Probe (WMAP) Observations: Cosmological Parameter Results. ApJS, 208:19. 3

Hoeft, M., Yepes, G., Gottlöber, S., and Springel, V. (2006). Dwarf galaxies in voids: suppressing star formation with photoheating. MNRAS, 371:401–414. 27

Hu, W. and White, M. (1997). A CMB polarization primer. NewA, 2:323–344. 13

Iliev, I. T., Mellema, G., Ahn, K., Shapiro, P. R., Mao, Y., and Pen, U.-L. (2014). Simulating cosmic reionization: how large a volume is large enough? MNRAS, 439:725–743. 64

Iliev, I. T., Mellema, G., Pen, U.-L., Bond, J. R., and Shapiro, P. R. (2008a). Current

models of the observable consequences of cosmic reionization and their detectability. MNRAS, 384:863–874. 21

Iliev, I. T., Mellema, G., Pen, U.-L., Merz, H., Shapiro, P. R., and Alvarez, M. A. (2006). Simulating cosmic reionization at large scales - I. The geometry of reionization. MNRAS, 369:1625–1638. 52

Iliev, I. T., Shapiro, P. R., Ferrara, A., and Martel, H. (2002). On the Direct Detectability of the Cosmic Dark Ages: 21 Centimeter Emission from Minihalos. ApJ, 572:L123–L126. 20, 21

Iliev, I. T., Shapiro, P. R., McDonald, P., Mellema, G., and Pen, U.-L. (2008b). The effect of the intergalactic environment on the observability of Lyα emitters during reionization. MNRAS, 391:63–83. 57

Iliev, I. T., Shapiro, P. R., and Raga, A. C. (2005). Minihalo photoevaporation during cosmic reionization: evaporation times and photon consumption rates. MNRAS, 361:405–414. 20, 27

Inoue, A. K., Iwata, I., and Deharveng, J.-M. (2006). The escape fraction of ionizing photons from galaxies at z = 0-6. MNRAS, 371:L1–L5. 17

Jaacks, J., Nagamine, K., and Choi, J. H. (2012). Duty cycle and the increasing star formation history of z $\geq$ 6 galaxies. MNRAS, 427:403–414. 26

Jeon, M., Pawlik, A. H., Bromm, V., and Milosavljević, M. (2014). Recovery from Population III supernova explosions and the onset of second-generation star formation. MNRAS, 444:3288–3300. 27

Kalberla, P. M. W. and Kerp, J. (2009). The Hi Distribution of the Milky Way. ARA&A, 47:27–61. 14

Kauffmann, G. (1996). Disc galaxies at z=0 and at high redshift: an explanation of the observed evolution of damped Lyalpha absorption systems. MNRAS, 281:475–486. 27, 127

Kauffmann, G., Heckman, T. M., White, S. D. M., Charlot, S., Tremonti, C., Brinchmann, J., Bruzual, G., Peng, E. W., Seibert, M., Bernardi, M., Blanton, M., Brinkmann, J., Castander, F., Csábai, I., Fukugita, M., Ivezic, Z., Munn, J. A., Nichol, R. C., Padmanabhan, N., Thakar, A. R., Weinberg, D. H., and York, D. (2003). Stellar masses

and star formation histories for 10^5 galaxies from the Sloan Digital Sky Survey. MNRAS, 341:33–53. 49

Kauffmann, G., White, S. D. M., and Guiderdoni, B. (1993). The Formation and Evolution of Galaxies Within Merging Dark Matter Haloes. MNRAS, 264:201. 5, 9, 26

Kauffmann, G., White, S. D. M., Heckman, T. M., Ménard, B., Brinchmann, J., Charlot, S., Tremonti, C., and Brinkmann, J. (2004). The environmental dependence of the relations between stellar mass, structure, star formation and nuclear activity in galaxies. MNRAS, 353:713–731. 31

Kennicutt, Jr., R. C. (1998). The Global Schmidt Law in Star-forming Galaxies. ApJ, 498:541–552. 126

Khaire, V., Srianand, R., Choudhury, T. R., and Gaikwad, P. (2016). The redshift evolution of escape fraction of hydrogen ionizing photons from galaxies. MNRAS, 457:4051–4062. 42

Klimentowski, J., Łokas, E. L., Knebe, A., Gottlöber, S., Martinez-Vaquero, L. A., Yepes, G., and Hoffman, Y. (2010). The grouping, merging and survival of subhaloes in the simulated Local Group. 402:1899–1910. 77

Klypin, A., Kravtsov, A. V., Valenzuela, O., and Prada, F. (1999). Where Are the Missing Galactic Satellites? ApJ, 522:82–92. 27

Klypin, A. A. and Shandarin, S. F. (1983). Three-dimensional numerical model of the formation of large-scale structure in the Universe. MNRAS, 204:891–907. 4

Knebe, A., Libeskind, N. I., Pearce, F., Behroozi, P., Casado, J., Dolag, K., Dominguez-Tenreiro, R., Elahi, P., Lux, H., Muldrew, S. I., and Onions, J. (2013a). Galaxies going MAD: the Galaxy-Finder Comparison Project. 428:2039–2052. 77

Knebe, A., Pearce, F. R., Lux, H., Ascasibar, Y., Behroozi, P., Casado, J., Moran, C. C., Diemand, J., Dolag, K., Dominguez-Tenreiro, R., Elahi, P., Falck, B., Gottlöber, S., Han, J., Klypin, A., Lukić, Z., Maciejewski, M., McBride, C. K., Merchán, M. E., Muldrew, S. I., Neyrinck, M., Onions, J., Planelles, S., Potter, D., Quilis, V., Rasera, Y., Ricker, P. M., Roy, F., Ruiz, A. N., Sgró, M. A., Springel, V., Stadel, J., Sutter, P. M., Tweed, D., and Zemp, M. (2013b). Structure finding in cosmological simulations: the state of affairs. MNRAS, 435:1618–1658. 5

Knollmann, S. R. and Knebe, A. (2009). AHF: Amiga's Halo Finder. ApJS, 182:608–624. 5

Knollmann, S. R. and Knebe, A. (2009). AHF: amiga's halo finder. 182:608. 70

Komatsu, E., Dunkley, J., Nolta, M. R., Bennett, C. L., Gold, B., Hinshaw, G., Jarosik, N., Larson, D., Limon, M., Page, L., Spergel, D. N., Halpern, M., Hill, R. S., Kogut, A., Meyer, S. S., Tucker, G. S., Weiland, J. L., Wollack, E., and Wright, E. L. (2009). Five-Year Wilkinson Microwave Anisotropy Probe Observations: Cosmological Interpretation. ApJS, 180:330–376. 3, 29

Komatsu, E., Smith, K. M., Dunkley, J., Bennett, C. L., Gold, B., Hinshaw, G., Jarosik, N., Larson, D., Nolta, M. R., Page, L., Spergel, D. N., Halpern, M., Hill, R. S., Kogut, A., Limon, M., Meyer, S. S., Odegard, N., Tucker, G. S., Weiland, J. L., Wollack, E., and Wright, E. L. (2011a). Seven-year Wilkinson Microwave Anisotropy Probe (WMAP) Observations: Cosmological Interpretation. ApJS, 192:18. 3

Komatsu, E., Smith, K. M., Dunkley, J., Bennett, C. L., Gold, B., Hinshaw, G., Jarosik, N., Larson, D., Nolta, M. R., Page, L., Spergel, D. N., Halpern, M., Hill, R. S., Kogut, A., Limon, M., Meyer, S. S., Odegard, N., Tucker, G. S., Weiland, J. L., Wollack, E., and Wright, E. L. (2011b). Seven-year Wilkinson Microwave Anisotropy Probe (WMAP) Observations: Cosmological Interpretation. 192:18. 70

Kuhlen, M., Krumholz, M. R., Madau, P., Smith, B. D., and Wise, J. (2012a). Dwarf Galaxy Formation with H_2-regulated Star Formation. ApJ, 749:36. 42

Kuhlen, M., Lisanti, M., and Spergel, D. N. (2012b). Direct detection of dark matter debris flows. Phys. Rev. D, 86(6):063505. 4

Lacey, C. and Cole, S. (1993). Merger rates in hierarchical models of galaxy formation. MNRAS, 262:627–649. 6, 26

Lacey, C. and Silk, J. (1991). Tidally triggered galaxy formation. I - Evolution of the galaxy luminosity function. ApJ, 381:14–32. 9

Lacey, C. G., Baugh, C. M., Frenk, C. S., and Benson, A. J. (2011). The evolution of Lyman-break galaxies in the cold dark matter model. MNRAS, 412:1828–1852. 27

Larson, R. B. (1974). Effects of supernovae on the early evolution of galaxies. MNRAS, 169:229–246. 127

Lee, J. and Yi, S. K. (2013). On the Assembly History of Stellar Components in Massive Galaxies. 766:38. 79

Lee, J., Yi, S. K., Elahi, P. J., Thomas, P. A., Pearce, F. R., Behroozi, P., Han, J., Helly, J., Jung, I., Knebe, A., Mao, Y.-Y., Onions, J., Rodriguez-Gomez, V., Schneider, A., Srisawat, C., and Tweed, D. (2014). Sussing merger trees: the impact of halo merger trees on galaxy properties in a semi-analytic model. MNRAS, 445:4197–4210. 31, 68

Lewis, A., Challinor, A., and Lasenby, A. (2000). Efficient Computation of Cosmic Microwave Background Anisotropies in Closed Friedmann-Robertson-Walker Models. ApJ, 538:473–476. 29

Libeskind, N. I., Knebe, A., Hoffman, Y., Gottlöber, S., and Yepes, G. (2011). Disentangling the dark matter halo from the stellar halo. 418:336–345. 77

Lidz, A., Oh, S. P., and Furlanetto, S. R. (2006). Have We Detected Patchy Reionization in Quasar Spectra? ApJ, 639:L47–L50. 25

Lidz, A., Zahn, O., Furlanetto, S. R., McQuinn, M., Hernquist, L., and Zaldarriaga, M. (2009). Probing Reionization with the 21 cm Galaxy Cross-Power Spectrum. ApJ, 690:252–266. 26

Loeb, A. and Barkana, R. (2001). The Reionization of the Universe by the First Stars and Quasars. ARA&A, 39:19–66. 11, 18, 21

Lorenzoni, S., Bunker, A. J., Wilkins, S. M., Caruana, J., Stanway, E. R., and Jarvis, M. J. (2013). Constraining the bright-end of the UV luminosity function for z $\sim$ 7-9 galaxies: results from CANDELS/GOODS-South. MNRAS, 429:150–158. 26

Lorenzoni, S., Bunker, A. J., Wilkins, S. M., Stanway, E. R., Jarvis, M. J., and Caruana, J. (2011). Star-forming galaxies at z $\sim$ 8-9 from Hubble Space Telescope/WFC3: implications for reionization. MNRAS, 414:1455–1466. 26

Ma, C.-P. and Bertschinger, E. (1995). Cosmological Perturbation Theory in the Synchronous and Conformal Newtonian Gauges. ApJ, 455:7. 13

Madau, P., Haardt, F., and Rees, M. J. (1999). Radiative Transfer in a Clumpy Universe. III. The Nature of Cosmological Ionizing Sources. ApJ, 514:648–659. 16

Madau, P., Rees, M. J., Volonteri, M., Haardt, F., and Oh, S. P. (2004). Early Reionization by Miniquasars. ApJ, 604:484–494. 18

Majumdar, S., Mellema, G., Datta, K. K., Jensen, H., Choudhury, T. R., Bharadwaj, S., and Friedrich, M. M. (2014). On the use of seminumerical simulations in predicting the 21-cm signal from the epoch of reionization. MNRAS, 443:2843–2861. 27, 34, 52, 63

Mapelli, M. and Ferrara, A. (2005). Background radiation from sterile neutrino decay and reionization. MNRAS, 364:2–12. 18

Maselli, A., Ferrara, A., and Ciardi, B. (2003). CRASH: a radiative transfer scheme. MNRAS, 345:379–394. 25

Mathis, J. S. (1983). Nebular dust and extinction in ionized nebulae. I - The Balmer decrement. ApJ, 267:119–125. 134

McLure, R. J., Dunlop, J. S., Bowler, R. A. A., Curtis-Lake, E., Schenker, M., Ellis, R. S., Robertson, B. E., Koekemoer, A. M., Rogers, A. B., Ono, Y., Ouchi, M., Charlot, S., Wild, V., Stark, D. P., Furlanetto, S. R., Cirasuolo, M., and Targett, T. A. (2013). A new multifield determination of the galaxy luminosity function at z = 7-9 incorporating the 2012 Hubble Ultra-Deep Field imaging. MNRAS, 432:2696–2716. 26

McLure, R. J., Dunlop, J. S., Cirasuolo, M., Koekemoer, A. M., Sabbi, E., Stark, D. P., Targett, T. A., and Ellis, R. S. (2010). Galaxies at z = 6-9 from the WFC3/IR imaging of the Hubble Ultra Deep Field. MNRAS, 403:960–983. 26

McLure, R. J., Dunlop, J. S., de Ravel, L., Cirasuolo, M., Ellis, R. S., Schenker, M., Robertson, B. E., Koekemoer, A. M., Stark, D. P., and Bowler, R. A. A. (2011). A robust sample of galaxies at redshifts 6.0<z<8.7: stellar populations, star formation rates and stellar masses. MNRAS, 418:2074–2105. 26, 38

McQuinn, M., Lidz, A., Zahn, O., Dutta, S., Hernquist, L., and Zaldarriaga, M. (2007). The morphology of HII regions during reionization. MNRAS, 377:1043–1063. 57

Meiksin, A. (2005). Constraints on the ionization sources of the high-redshift intergalactic medium. MNRAS, 356:596–606. 16

Meiksin, A. A. (2009). The physics of the intergalactic medium. *Reviews of Modern Physics*, 81:1405–1469. 18

Mellema, G., Iliev, I. T., Alvarez, M. A., and Shapiro, P. R. (2006). C ^{2}-ray: A new method for photon-conserving transport of ionizing radiation. NewA, 11:374–395. 25, 28

Mesinger, A. and Furlanetto, S. (2007). Efficient Simulations of Early Structure Formation and Reionization. ApJ, 669:663–675. 26, 34

Mesinger, A. and Furlanetto, S. (2009). The inhomogeneous ionizing background following reionization. MNRAS, 400:1461–1471. 25

Mitra, S., Choudhury, T. R., and Ferrara, A. (2015). Cosmic reionization after Planck. MNRAS, 454:L76–L80. 42

Moore, B., Ghigna, S., Governato, F., Lake, G., Quinn, T., Stadel, J., and Tozzi, P. (1999). Dark Matter Substructure within Galactic Halos. ApJ, 524:L19–L22. 27

Mortlock, D. J., Warren, S. J., Venemans, B. P., Patel, M., Hewett, P. C., McMahon, R. G., Simpson, C., Theuns, T., Gonzáles-Solares, E. A., Adamson, A., Dye, S., Hambly, N. C., Hirst, P., Irwin, M. J., Kuiper, E., Lawrence, A., and Röttgering, H. J. A. (2011). A luminous quasar at a redshift of z = 7.085. Nature, 474:616–619. 16

Mutch, S. J., Geil, P. M., Poole, G. B., Angel, P. W., Duffy, A. R., Mesinger, A., and Wyithe, J. S. B. (2015). Dark-ages reionization and galaxy formation simulation III: Modelling galaxy formation and the Epoch of Reionization. *ArXiv e-prints*. 11, 27, 48, 49, 52, 63, 64

Norberg, P., Baugh, C. M., Gaztañaga, E., and Croton, D. J. (2009). Statistical analysis of galaxy surveys - I. Robust error estimation for two-point clustering statistics. MNRAS, 396:19–38. 54

Norman, M. L., Reynolds, D. R., So, G. C., Harkness, R. P., and Wise, J. H. (2015). Fully Coupled Simulation of Cosmic Reionization. I. Numerical Methods and Tests. ApJS, 216:16. 27

Ocvirk, P., Gillet, N., Shapiro, P. R., Aubert, D., Iliev, I. T., Teyssier, R., Yepes, G., Choi, J.-H., Sullivan, D., Knebe, A., Gottlöber, S., D'Aloisio, A., Park, H., Hoffman, Y., and Stranex, T. (2016). Cosmic Dawn (CoDa): the First Radiation-Hydrodynamics Simulation of Reionization and Galaxy Formation in the Local Universe. MNRAS. 27

Oesch, P. A., Bouwens, R. J., Illingworth, G. D., Carollo, C. M., Franx, M., Labbé, I., Magee, D., Stiavelli, M., Trenti, M., and van Dokkum, P. G. (2010). z ˜ 7 Galaxies in the HUDF: First Epoch WFC3/IR Results. ApJ, 709:L16–L20. 26, 38

Okamoto, T., Gao, L., and Theuns, T. (2008). Mass loss of galaxies due to an ultraviolet background. MNRAS, 390:920–928. 11, 21, 22, 27, 32, 36, 37, 42, 43, 44, 45

Onions, J., Ascasibar, Y., Behroozi, P., Casado, J., Elahi, P., Han, J., Knebe, A., Lux, H., Merchán, M. E., Muldrew, S. I., Neyrinck, M., Old, L., Pearce, F. R., Potter, D., Ruiz, A. N., Sgró, M. A., Tweed, D., and Yue, T. (2013). Subhaloes gone Notts: spin across subhaloes and finders. MNRAS, 429:2739–2747. 6

Onions, J., Knebe, A., Pearce, F. R., Muldrew, S. I., Lux, H., Knollmann, S. R., Ascasibar, Y., Behroozi, P., Elahi, P., Han, J., Maciejewski, M., Merchán, M. E., Neyrinck, M., Ruiz, A. N., Sgró, M. A., Springel, V., and Tweed, D. (2012). Subhaloes going Notts: the subhalo-finder comparison project. MNRAS, 423:1200–1214. 6

Ouchi, M., Ono, Y., Egami, E., Saito, T., Oguri, M., McCarthy, P. J., Farrah, D., Kashikawa, N., Momcheva, I., Shimasaku, K., Nakanishi, K., Furusawa, H., Akiyama, M., Dunlop, J. S., Mortier, A. M. J., Okamura, S., Hayashi, M., Cirasuolo, M., Dressler, A., Iye, M., Jarvis, M. J., Kodama, T., Martin, C. L., McLure, R. J., Ohta, K., Yamada, T., and Yoshida, M. (2009). Discovery of a Giant Lyα Emitter Near the Reionization Epoch. ApJ, 696:1164–1175. 16

Park, H., Shapiro, P. R., Komatsu, E., Iliev, I. T., Ahn, K., and Mellema, G. (2013). The Kinetic Sunyaev-Zel'dovich Effect as a Probe of the Physics of Cosmic Reionization: The Effect of Self-regulated Reionization. ApJ, 769:93. 16

Parsons, A. R., Liu, A., Aguirre, J. E., Ali, Z. S., Bradley, R. F., Carilli, C. L., DeBoer, D. R., Dexter, M. R., Gugliucci, N. E., Jacobs, D. C., Klima, P., MacMahon, D. H. E., Manley, J. R., Moore, D. F., Pober, J. C., Stefan, I. I., and Walbrugh, W. P. (2014). New Limits on 21 cm Epoch of Reionization from PAPER-32 Consistent with an X-Ray Heated Intergalactic Medium at z = 7.7. ApJ, 788:106. 15

Pawlik, A. H., Milosavljević, M., and Bromm, V. (2013). The First Galaxies: Assembly under Radiative Feedback from the First Stars. ApJ, 767:59. 27

Pawlik, A. H., Schaye, J., and Dalla Vecchia, C. (2015). Spatially adaptive radiation-hydrodynamical simulations of galaxy formation during cosmological reionization. MNRAS, 451:1586–1605. 27

Pawlik, A. H., Schaye, J., and van Scherpenzeel, E. (2009). Keeping the Universe ionized: photoheating and the clumping factor of the high-redshift intergalactic medium. MNRAS, 394:1812–1824. 35

Peacock, J. A., Cole, S., Norberg, P., Baugh, C. M., Bland-Hawthorn, J., Bridges, T., Cannon, R. D., Colless, M., Collins, C., Couch, W., Dalton, G., Deeley, K., De Propris,

R., Driver, S. P., Efstathiou, G., Ellis, R. S., Frenk, C. S., Glazebrook, K., Jackson, C., Lahav, O., Lewis, I., Lumsden, S., Maddox, S., Percival, W. J., Peterson, B. A., Price, I., Sutherland, W., and Taylor, K. (2001). A measurement of the cosmological mass density from clustering in the 2dF Galaxy Redshift Survey. Nature, 410:169–173. 3

Peebles, P. J. E. (1982). Large-scale background temperature and mass fluctuations due to scale-invariant primeval perturbations. ApJ, 263:L1–L5. 3

Peebles, P. J. E. and Yu, J. T. (1970). Primeval Adiabatic Perturbation in an Expanding Universe. ApJ, 162:815. 13

Pen, U.-L., Chang, T.-C., Hirata, C. M., Peterson, J. B., Roy, J., Gupta, Y., Odegova, J., and Sigurdson, K. (2009). The GMRT EoR experiment: limits on polarized sky brightness at 150 MHz. MNRAS, 399:181–194. 15

Perlmutter, S., Aldering, G., Goldhaber, G., Knop, R. A., Nugent, P., Castro, P. G., Deustua, S., Fabbro, S., Goobar, A., Groom, D. E., Hook, I. M., Kim, A. G., Kim, M. Y., Lee, J. C., Nunes, N. J., Pain, R., Pennypacker, C. R., Quimby, R., Lidman, C., Ellis, R. S., Irwin, M., McMahon, R. G., Ruiz-Lapuente, P., Walton, N., Schaefer, B., Boyle, B. J., Filippenko, A. V., Matheson, T., Fruchter, A. S., Panagia, N., Newberg, H. J. M., Couch, W. J., and Project, T. S. C. (1999). Measurements of Ω and Λ from 42 High-Redshift Supernovae. ApJ, 517:565–586. 2

Planck Collaboration, Ade, P. A. R., Aghanim, N., Armitage-Caplan, C., Arnaud, M., Ashdown, M., Atrio-Barandela, F., Aumont, J., Baccigalupi, C., Banday, A. J., and et al. (2014). Planck 2013 results. XVI. Cosmological parameters. A&A, 571:A16. 3

Planck Collaboration, Ade, P. A. R., Aghanim, N., Arnaud, M., Ashdown, M., Aumont, J., Baccigalupi, C., Banday, A. J., Barreiro, R. B., Bartlett, J. G., and et al. (2015). Planck 2015 results. XIII. Cosmological parameters. *ArXiv e-prints*. 10, 11, 3, 13, 41, 43, 63

Press, W. H. and Schechter, P. (1974). Formation of Galaxies and Clusters of Galaxies by Self-Similar Gravitational Condensation. ApJ, 187:425–438. 3, 5, 9

Pritchard, J. R. and Loeb, A. (2012). 21 cm cosmology in the 21st century. *Reports on Progress in Physics*, 75(8):086901. 10, 14, 15

Putman, M. E., Bland-Hawthorn, J., Veilleux, S., Gibson, B. K., Freeman, K. C., and

Maloney, P. R. (2003). Hα Emission from High-Velocity Clouds and Their Distances. ApJ, 597:948–956. 17

Raičević, M., Theuns, T., and Lacey, C. (2011). The galaxies that reionized the Universe. MNRAS, 410:775–787. 27

Razoumov, A. O. and Cardall, C. Y. (2005). Fully threaded transport engine: new method for multi-scale radiative transfer. MNRAS, 362:1413–1417. 25

Reed, D. S., Smith, R. E., Potter, D., Schneider, A., Stadel, J., and Moore, B. (2013). Towards an accurate mass function for precision cosmology. MNRAS, 431:1866–1882. 4

Ricotti, M. and Ostriker, J. P. (2004). Reionization, chemical enrichment and seed black holes from the first stars: is Population III important? MNRAS, 350:539–551. 18

Ricotti, M., Ostriker, J. P., and Gnedin, N. Y. (2005). X-ray pre-ionization powered by accretion on the first black holes - II. Cosmological simulations and observational signatures. MNRAS, 357:207–219. 18

Rijkhorst, E.-J., Plewa, T., Dubey, A., and Mellema, G. (2006). Hybrid characteristics: 3D radiative transfer for parallel adaptive mesh refinement hydrodynamics. A&A, 452:907–920. 25

Ritzerveld, J., Icke, V., and Rijkhorst, E.-J. (2003). Triangulating Radiation: Radiative Transfer on Unstructured Grids. *ArXiv Astrophysics e-prints*. 25

Robertson, B. E., Ellis, R. S., Dunlop, J. S., McLure, R. J., and Stark, D. P. (2010). Early star-forming galaxies and the reionization of the Universe. Nature, 468:49–55. 18

Robertson, B. E., Ellis, R. S., Furlanetto, S. R., and Dunlop, J. S. (2015). Cosmic Reionization and Early Star-forming Galaxies: A Joint Analysis of New Constraints from Planck and the Hubble Space Telescope. ApJ, 802:L19. 17

Robertson, B. E., Furlanetto, S. R., Schneider, E., Charlot, S., Ellis, R. S., Stark, D. P., McLure, R. J., Dunlop, J. S., Koekemoer, A., Schenker, M. A., Ouchi, M., Ono, Y., Curtis-Lake, E., Rogers, A. B., Bowler, R. A. A., and Cirasuolo, M. (2013). New Constraints on Cosmic Reionization from the 2012 Hubble Ultra Deep Field Campaign. ApJ, 768:71. 26

Rosdahl, J., Schaye, J., Teyssier, R., and Agertz, O. (2015). Galaxies that shine: radiation-hydrodynamical simulations of disc galaxies. MNRAS, 451:34–58. 27

Roukema, B. F., Quinn, P. J., Peterson, B. A., and Rocca-Volmerange, B. (1997). Merging history trees of dark matter haloes - A tool for exploring galaxy formation models. MNRAS, 292:835. 6

Salvaterra, R., Ferrara, A., and Dayal, P. (2011). Simulating high-redshift galaxies. MNRAS, 414:847–859. 26

Santos, M. G., Ferramacho, L., Silva, M. B., Amblard, A., and Cooray, A. (2010). Fast large volume simulations of the 21-cm signal from the reionization and pre-reionization epochs. MNRAS, 406:2421–2432. 26, 34

Schneider, R., Salvaterra, R., Ferrara, A., and Ciardi, B. (2006). Constraints on the initial mass function of the first stars. MNRAS, 369:825–834. 18

Shapiro, P. R. and Giroux, M. L. (1987). Cosmological H II regions and the photoionization of the intergalactic medium. ApJ, 321:L107–L112. 16

Shapiro, P. R., Giroux, M. L., and Babul, A. (1994). Reionization in a cold dark matter universe: The feedback of galaxy formation on the intergalactic medium. ApJ, 427:25–50. 21, 27

Shapiro, P. R., Iliev, I. T., and Raga, A. C. (2004). Photoevaporation of cosmological minihaloes during reionization. MNRAS, 348:753–782. 21, 27

Shapiro, P. R., Iliev, I. T., Raga, A. C., and Martel, H. (2003). Photoevaporation of Minihalos during Reionization. In Holt, S. H. and Reynolds, C. S., editors, *The Emergence of Cosmic Structure*, volume 666 of *American Institute of Physics Conference Series*, pages 89–92. 20

Shapley, A. E., Steidel, C. C., Pettini, M., Adelberger, K. L., and Erb, D. K. (2006). The Direct Detection of Lyman Continuum Emission from Star-forming Galaxies at z˜3. ApJ, 651:688–703. 17

Smit, R., Bouwens, R. J., Franx, M., Illingworth, G. D., Labbé, I., Oesch, P. A., and van Dokkum, P. G. (2012). The Star Formation Rate Function for Redshift z ˜ 4-7 Galaxies: Evidence for a Uniform Buildup of Star-forming Galaxies during the First 3 Gyr of Cosmic Time. ApJ, 756:14. 11, 47, 48

Smoot, G. F., Bennett, C. L., Kogut, A., Wright, E. L., Aymon, J., Boggess, N. W., Cheng, E. S., de Amici, G., Gulkis, S., Hauser, M. G., Hinshaw, G., Jackson, P. D., Janssen, M., Kaita, E., Kelsall, T., Keegstra, P., Lineweaver, C., Loewenstein, K., Lubin, P., Mather, J., Meyer, S. S., Moseley, S. H., Murdock, T., Rokke, L., Silverberg, R. F., Tenorio, L., Weiss, R., and Wilkinson, D. T. (1992). Structure in the COBE differential microwave radiometer first-year maps. ApJ, 396:L1–L5. 2

So, G. C., Norman, M. L., Reynolds, D. R., and Wise, J. H. (2014). Fully Coupled Simulation of Cosmic Reionization. II. Recombinations, Clumping Factors, and the Photon Budget for Reionization. ApJ, 789:149. 27

Sobral, D., Matthee, J., Darvish, B., Schaerer, D., Mobasher, B., Röttgering, H. J. A., Santos, S., and Hemmati, S. (2015). Evidence for PopIII-like Stellar Populations in the Most Luminous Lyman-α Emitters at the Epoch of Reionization: Spectroscopic Confirmation. ApJ, 808:139. 18

Somerville, R. S., Hopkins, P. F., Cox, T. J., Robertson, B. E., and Hernquist, L. (2008). A semi-analytic model for the co-evolution of galaxies, black holes and active galactic nuclei. MNRAS, 391:481–506. 9

Somerville, R. S. and Primack, J. R. (1999). Semi-analytic modelling of galaxy formation: the local Universe. MNRAS, 310:1087–1110. 9

Somerville, R. S., Primack, J. R., and Faber, S. M. (2001). The nature of high-redshift galaxies. MNRAS, 320:504–528. 9, 132

Song, H., Park, C., Lietzen, H., and Einasto, M. (2016). Quasars as a Tracer of Large-scale Structures in the Distant Universe. ApJ, 827:104. 11, 47, 48, 49, 63

Spergel, D. N., Bean, R., Doré, O., Nolta, M. R., Bennett, C. L., Dunkley, J., Hinshaw, G., Jarosik, N., Komatsu, E., Page, L., Peiris, H. V., Verde, L., Halpern, M., Hill, R. S., Kogut, A., Limon, M., Meyer, S. S., Odegard, N., Tucker, G. S., Weiland, J. L., Wollack, E., and Wright, E. L. (2007). Three-Year Wilkinson Microwave Anisotropy Probe (WMAP) Observations: Implications for Cosmology. ApJS, 170:377–408. 3

Spergel, D. N., Verde, L., Peiris, H. V., Komatsu, E., Nolta, M. R., Bennett, C. L., Halpern, M., Hinshaw, G., Jarosik, N., Kogut, A., Limon, M., Meyer, S. S., Page, L., Tucker, G. S., Weiland, J. L., Wollack, E., and Wright, E. L. (2003). First-Year Wilkinson Microwave Anisotropy Probe (WMAP) Observations: Determination of Cosmological Parameters. ApJS, 148:175–194. 3

Springel, V. (2005). The cosmological simulation code gadget2. 364(4):1105–1134. 70

Springel, V., Di Matteo, T., and Hernquist, L. (2005a). Modelling feedback from stars and black holes in galaxy mergers. MNRAS, 361:776–794. 31

Springel, V., White, S. D. M., Jenkins, A., Frenk, C. S., Yoshida, N., Gao, L., Navarro, J., Thacker, R., Croton, D., Helly, J., Peacock, J. A., Cole, S., Thomas, P., Couchman, H., Evrard, A., Colberg, J., and Pearce, F. (2005b). Simulations of the formation, evolution and clustering of galaxies and quasars. Nature, 435:629–636. 10, 30

Springel, V., White, S. D. M., Jenkins, A., Frenk, C. S., Yoshida, N., Gao, L., Navarro, J., Thacker, R., Croton, D., Helly, J., Peacock, J. A., Cole, S., Thomas, P., Couchman, H., Evrard, A., Colberg, J., and Pearce, F. (2005c). Simulations of the formation, evolution and clustering of galaxies and quasars. 435:629–636. 76

Springel, V., White, S. D. M., Tormen, G., and Kauffmann, G. (2001a). Populating a cluster of galaxies - I. Results at [formmu2]z=0. MNRAS, 328:726–750. 5, 6, 30, 31, 125

Springel, V., Yoshida, N., and White, S. D. M. (2001b). GADGET: a code for collisionless and gasdynamical cosmological simulations. NewA, 6:79–117. 4

Srbinovsky, J. A. and Wyithe, J. S. B. (2007). Constraining the quasar contribution to the reionization of cosmic hydrogen. MNRAS, 374:627–633. 16

Srisawat, C., Knebe, A., Pearce, F. R., Schneider, A., Thomas, P. A., Behroozi, P., Dolag, K., Elahi, P. J., Han, J., Helly, J., Jing, Y., Jung, I., Lee, J., Mao, Y.-Y., Onions, J., Rodriguez-Gomez, V., Tweed, D., and Yi, S. K. (2013). Sussing Merger Trees: The Merger Trees Comparison Project. MNRAS, 436:150–162. 31, 32, 67, 68

Sugiyama, N., Zaroubi, S., and Silk, J. (2004). Isocurvature fluctuations induce early star formation. MNRAS, 354:543–548. 18

Sunyaev, R. A. and Zeldovich, Y. B. (1972). The Observations of Relic Radiation as a Test of the Nature of X-Ray Radiation from the Clusters of Galaxies. *Comments on Astrophysics and Space Physics*, 4:173. 13

Sutherland, R. S. and Dopita, M. A. (1993). Cooling functions for low-density astrophysical plasmas. ApJS, 88:253–327. 45, 64, 125

Tingay, S. J., Goeke, R., Bowman, J. D., Emrich, D., Ord, S. M., Mitchell, D. A., Morales, M. F., Booler, T., Crosse, B., Wayth, R. B., Lonsdale, C. J., Tremblay, S., Pallot, D.,

Colegate, T., Wicenec, A., Kudryavtseva, N., Arcus, W., Barnes, D., Bernardi, G., Briggs, F., Burns, S., Bunton, J. D., Cappallo, R. J., Corey, B. E., Deshpande, A., Desouza, L., Gaensler, B. M., Greenhill, L. J., Hall, P. J., Hazelton, B. J., Herne, D., Hewitt, J. N., Johnston-Hollitt, M., Kaplan, D. L., Kasper, J. C., Kincaid, B. B., Koenig, R., Kratzenberg, E., Lynch, M. J., Mckinley, B., Mcwhirter, S. R., Morgan, E., Oberoi, D., Pathikulangara, J., Prabu, T., Remillard, R. A., Rogers, A. E. E., Roshi, A., Salah, J. E., Sault, R. J., Udaya-Shankar, N., Schlagenhaufer, F., Srivani, K. S., Stevens, J., Subrahmanyan, R., Waterson, M., Webster, R. L., Whitney, A. R., Williams, A., Williams, C. L., and Wyithe, J. S. B. (2013). The Murchison Widefield Array: The Square Kilometre Array Precursor at Low Radio Frequencies. PASA, 30:7. 15

Tonini, C., Maraston, C., Devriendt, J., Thomas, D., and Silk, J. (2009). The impact of thermally pulsing asymptotic giant branch stars on hierarchical galaxy formation models. MNRAS, 396:L36–L40. 133

Tonini, C., Maraston, C., Thomas, D., Devriendt, J., and Silk, J. (2010). Hierarchical models of high-redshift galaxies with thermally pulsing asymptotic giant branch stars: comparison with observations. MNRAS, 403:1749–1758. 133

Toomre, A. (1964). On the gravitational stability of a disk of stars. ApJ, 139:1217–1238. 127

Topping, M. W. and Shull, J. M. (2015). The Efficiency of Stellar Reionization: Effects of Rotation, Metallicity, and Initial Mass Function. ApJ, 800:97. 42

Trotta, R. (2008). Bayes in the sky: Bayesian inference and model selection in cosmology. *Contemporary Physics*, 49:71–104. 10

Tumlinson, J., Shull, J. M., and Venkatesan, A. (2002). Cosmological Reionization by the First Stars: Evolving Spectra of Population III. In Crowther, P., editor, *Hot Star Workshop III: The Earliest Phases of Massive Star Birth*, volume 267 of *Astronomical Society of the Pacific Conference Series*, page 433. 18

Tweed, D., Devriendt, J., Blaizot, J., Colombi, S., and Slyz, A. (2009). Building merger trees from cosmological N-body simulations. Towards improving galaxy formation models using subhaloes. 506:647–660. 78

Venkatesan, A., Giroux, M. L., and Shull, J. M. (2001). Heating and Ionization of the Intergalactic Medium by an Early X-Ray Background. ApJ, 563:1–8. 18

Wang, L., Dutton, A. A., Stinson, G. S., Macciò, A. V., Gutcke, T., and Kang, X. (2016). NIHAO VII: Predictions for the galactic baryon budget in dwarf to Milky Way mass haloes. *ArXiv e-prints*. 31

Warmels, R. H. (1988). The HI properties of spiral galaxies in the Virgo Cluster. III - The HI surface density distribution in 36 galaxies. A&AS, 72:427–447. 14

Whalen, D. and Norman, M. L. (2006). A Multistep Algorithm for the Radiation Hydrodynamical Transport of Cosmological Ionization Fronts and Ionized Flows. ApJS, 162:281–303. 25

White, R. L., Becker, R. H., Fan, X., and Strauss, M. A. (2003). Probing the Ionization State of the Universe at z ¿ 6. AJ, 126:1–14. 12

White, S. D. M. and Frenk, C. S. (1991). Galaxy formation through hierarchical clustering. ApJ, 379:52–79. 9, 31, 124, 125

White, S. D. M. and Rees, M. J. (1978). Core condensation in heavy halos - A two-stage theory for galaxy formation and clustering. MNRAS, 183:341–358. 3, 8, 127

Wilkins, S. M., Bunker, A., Coulton, W., Croft, R., Matteo, T. D., Khandai, N., and Feng, Y. (2013). Interpreting the observed UV continuum slopes of high-redshift galaxies. MNRAS, 430:2885–2890. 47

Wilkins, S. M., Bunker, A. J., Ellis, R. S., Stark, D., Stanway, E. R., Chiu, K., Lorenzoni, S., and Jarvis, M. J. (2010). Probing ˜L$_*$ Lyman-break galaxies at z ˜ 7 in GOODS-South with WFC3 on Hubble Space Telescope. MNRAS, 403:938–944. 26

Wilkins, S. M., Bunker, A. J., Lorenzoni, S., and Caruana, J. (2011a). New star-forming galaxies at z $\sim$ 7 from Wide Field Camera Three imaging. MNRAS, 411:23–36. 26

Wilkins, S. M., Bunker, A. J., Stanway, E., Lorenzoni, S., and Caruana, J. (2011b). The ultraviolet properties of star-forming galaxies - I. HST WFC3 observations of very high redshift galaxies. MNRAS, 417:717–729. 26

Wise, J. H., Demchenko, V. G., Halicek, M. T., Norman, M. L., Turk, M. J., Abel, T., and Smith, B. D. (2014). The birth of a galaxy - III. Propelling reionization with the faintest galaxies. MNRAS, 442:2560–2579. 27

Wouthuysen, S. A. (1952). On the excitation mechanism of the 21-cm (radio-frequency) interstellar hydrogen emission line. AJ, 57:31–32. 14

Wyithe, J. S. B. and Loeb, A. (2003). Was the Universe Reionized by Massive Metal-free Stars? ApJ, 588:L69–L72. 49

Wyithe, J. S. B. and Loeb, A. (2004). A characteristic size of ˜10Mpc for the ionized bubbles at the end of cosmic reionization. Nature, 432:194–196. 25

Wyithe, J. S. B. and Loeb, A. (2013). A suppressed contribution of low-mass galaxies to reionization due to supernova feedback. MNRAS, 428:2741–2754. 25, 49

Zahn, O., Lidz, A., McQuinn, M., Dutta, S., Hernquist, L., Zaldarriaga, M., and Furlanetto, S. R. (2007). Simulations and Analytic Calculations of Bubble Growth during Hydrogen Reionization. ApJ, 654:12–26. 26, 34, 52, 57

Zaldarriaga, M., Furlanetto, S. R., and Hernquist, L. (2004). 21 Centimeter Fluctuations from Cosmic Gas at High Redshifts. ApJ, 608:622–635. 13

Zaldarriaga, M., Scoccimarro, R., and Hui, L. (2003). Inferring the Linear Power Spectrum from the Lyα Forest. ApJ, 590:1–7. 16

Zaroubi, S. (2013). The Epoch of Reionization. In Wiklind, T., Mobasher, B., and Bromm, V., editors, *Astrophysics and Space Science Library*, volume 396 of *Astrophysics and Space Science Library*, page 45. 16

Zhou, J., Guo, Q., Liu, G.-C., Yue, B., Xu, Y.-D., and Chen, X.-L. (2013). Semi-numerical simulation of reionization with semi-analytical modeling of galaxy formation. *Research in Astronomy and Astrophysics*, 13:373–386. 27

Appendix A

The Munich semi-analytic model of galaxy formation L-Galaxies

This chapter is a brief summary of description of L-Galaxies described in the supplementary material of Henriques et al. (2015). However, we strongly encourage the readers to the referred document for more details.

A.1 Baryonic infall

We assume that collapsed dark matter haloes (FoF groups in this work) always have a mass of associated baryons given by the universal baryon fraction, $f_b = \Omega_b/\Omega_m$ (White and Frenk, 1991). This value specifies the maximum baryons a halo can possess. At any time, the matter which has not previously accreted into any objects will be added into the central objects in the form of primordial gas. However, some halos might not be able to accrete the total amount of baryons due to many factors, especially the ionising UV radiation which can heat the IGM and increase the local Jeans mass (Dijkstra et al., 2004b). We parametrize the reduction of infall gas in terms of a non-negative baryon fraction modifier, $f_{\rm mod}$, which specifies the total baryon mass a halo can capture in its lifetime. We model the infall mass,

$$m_{\rm infall} = f_{\rm mod} f_b M_{\rm 200c} - \sum_{i=0}^{N_{\rm gal}-1} (m^i_* + m^i_{\rm cold} + m^i_{\rm hot} + m^i_{\rm ejected} + m^i_{\rm BH}), \tag{A.1}$$

where $0 \leq f_{\rm mod} \leq 1$ and $N_{\rm gal}$ is the total number of galaxies inside the virial radius, $R_{\rm 200c}$, of the halo. m_*, $m_{\rm cold}$, $m_{\rm hot}$, $m_{\rm ejected}$, and $m_{\rm BH}$ are the total mass of stars, cold gas disc, hot atmosphere, ejected phase gas, and super massive black hole respectively. In

the standard model in L-GALAXIES, $m_{\rm infall}$ is always equal or greater than 0. If the right hand side of Equation 4.2 is negative, $m_{\rm infall}$ is set to 0.

A.2 Gas cooling

We follow White and Frenk (1991) and Springel et al. (2001a) that the infall gas which is accreted into the halo will be shock-heated up to the virial temperature of the host halo and added to the hydrostatic hot gas reservoir. This gas will be assumed, in the quasi-static regime, to be cooled down. The cooling time at a given radius, r, is then given by

$$t_{\rm cool}(r) = \frac{3\mu m_{\rm H} k T_{\rm vir}}{2\rho_{\rm hot}(r)\Lambda(T_{\rm hot}, Z_{\rm hot})} \tag{A.2}$$

where $\mu m_{\rm H}$ is the mean particle mass, k is the Boltzmann constant, $\rho_{\rm hot}(r)$ is the hot gas density, and $Z_{\rm hot}$ is the hot gas metallicity. $T_{\rm hot}$ is the temperature of the hot gas which is assumed to be the virial temperature of the halo given by $T_{\rm vir} = T_{\rm 200c} = 35.9(v_{\rm 200c}/{\rm kms}^{-1})^2$ K (for subhaloes we use this temperature as estimated at infall). $\Lambda(T_{\rm hot}, Z_{\rm hot})$ is the equilibrium cooling function for collisional processes, which depends both on the metallicity and temperature of the gas but ignores radiative ionization effects (Sutherland and Dopita, 1993). The hot gas density as a function of radius for an isothermal model is given by

$$\rho_{\rm hot}(r) = \frac{m_{\rm hot}}{4\pi R_{\rm 200c} r^2}. \tag{A.3}$$

The cooling time is assumed to be equal to the halo dynamical time

$$r_{\rm cool} = \left[\frac{t_{\rm dyn,h} m_{\rm hot} \Lambda\left(T_{\rm hot}, Z_{\rm hot}\right)}{6\pi\mu m_{\rm H} k T_{\rm vir} R_{\rm 200c}}\right]^{\frac{1}{2}}, \tag{A.4}$$

where $t_{\rm dyn,h}$ is the halo dynamic time defined as $R_{\rm vir}/v_{\rm vir} = R_{\rm 200c}/v_{\rm 200c} = 0.1/H(z)$ (De Lucia et al., 2004a). Thus, the cooling rate is given by Guo et al. (2011) as

$$\dot{m}_{\rm cool} = \begin{cases} m_{\rm hot}\frac{r_{\rm cool}}{R_{\rm 200c}}\frac{1}{t_{\rm dyn,h}}, & \text{if } r_{\rm cool} < R_{\rm 200c} \\ \frac{m_{\rm hot}}{t_{\rm dyn,h}}, & \text{if } r_{\rm cool} > R_{\rm 200c}. \end{cases} \tag{A.5}$$

A.3 Disc formation

We follow the simple model in Guo et al. (2011) such that when the material accretes into a halo, its dark matter and baryonic components are expected to have the same

angular momenta. When some of this gas is added to the central galaxy, the remaining angular momentum will determine the radius at which it settles inside the galactic disc. The properties of the cold gas and stellar discs are modelled separately. The angular momentum of the cold gas disc is given by

$$\Delta \mathbf{J}_{\rm gas} = \Delta \mathbf{J}_{\rm gas,cooling} + \Delta \mathbf{J}_{\rm gas,SF} + \Delta \mathbf{J}_{\rm gas,acc}, \tag{A.6}$$

where $\Delta \mathbf{J}_{\rm gas,cooling}, \Delta \mathbf{J}_{\rm gas,SF}$, and $\Delta \mathbf{J}_{\rm gas,acc}$ are the changes as the results of cooled gas, star formation, and accreted gas from minor mergers respectively. This can be expressed as

$$\Delta \mathbf{J}_{\rm gas} = \frac{\mathbf{J}_{\rm DM}}{M_{\rm DM}} \dot{m}_{\rm cool} \Delta t - \frac{\mathbf{J}_{\rm gas}}{m_{\rm gas}} \left((1 - R_{rec})\, \dot{m}_* \Delta t + \Delta m_{\rm reheat} \right) + \frac{\mathbf{J}_{\rm DM}}{M_{\rm DM}} m_{\rm sat,gas}, \tag{A.7}$$

where $\mathbf{J}_{\rm gas}$ and $M_{\rm DM}$ are the angular momentum and total mass of the FoF group respectively. Δt is the time interval considered, $(1 - R_{rec})\dot{m}_*$ is the formation rate of long lived stars (see Equation A.13), $\Delta m_{\rm reheat}$ is the cold gas reheated into hot atmosphere due to the star formation described in Section A.5, $\dot{m}_{\rm cool}$ is the cooling rate from Equation A.5, and $m_{\rm sat,gas}$ is the cold gas mass from merging satellites.

For stellar and gas discs, we assume them to be thin, in centrifugal equilibrium and have exponential surface density profiles:

$$\Sigma_{\rm gas}(R) = \Sigma_{\rm gas,0} \exp(-R/R_{\rm gas}) \tag{A.8}$$

and

$$\Sigma_*(R) = \Sigma_{*,0} \exp(-R/R_*), \tag{A.9}$$

where $\Sigma_{\rm gas,0} = m_{\rm gas}/2\pi R_{\rm gas}^2$ and $\Sigma_{\rm gas,0} = m_*/2\pi R_*^2$. By assuming flat circular velocity curve, the corresponding scale-lengths are given by

$$R_{\rm gas} = \frac{J_{\rm gas}}{2 m_{\rm gas} v_{\rm max}} \tag{A.10}$$

and

$$R_* = \frac{J_*}{2 m_{*,\rm d} v_{\rm max}}, \tag{A.11}$$

where $m_{\rm gas}$ and $m_{*,\rm d}$ are the total mass of the gas and stellar discs respectively.

A.4 Star formation

We assume stars to form from cold gas in the disc according to the simplified version of the empirical relation in Kennicutt (1998). Stars form efficiently only in the regions

where the surface mass density is greater than a critical surface density (Toomre, 1964). In this work, we follow Guo et al. (2011) to assume a flat rotation curve and a gas velocity dispersion of 6 km/s. Then, we use the model suggested by Kauffmann (1996), Croton et al. (2006) and Henriques et al. (2015) to estimate the critical mass:

$$m_{\rm crit} = m_{\rm crit,0} \left(\frac{v_{\rm max}}{200 \ {\rm kms}^{-1}} \right) \left(\frac{R_{\rm gas,d}}{10 \ {\rm kpc}} \right) {\rm M}_{\odot}, \tag{A.12}$$

where $m_{\rm crit,0}$ is a free parameter. The amount of the cold disc gas which is used to form long-lived per unit time is

$$\dot{m}_{*} = \alpha_{\rm SF} \frac{(m_{\rm cold} - m_{\rm crit})}{t_{\rm dyn,disk}} \tag{A.13}$$

where $t_{\rm dyn,disk} = R_{*}/v_{\rm max}$ is the characteristic dynamical time of the disc and $\alpha_{\rm SF}$ is the star formation coefficient.

A.5 Supernova feedback

In the final phase of massive stars, a large number of them explode as supernovae. The energy released by these supernovae drastically affects the surrounding inter-stellar medium (ISM). This causes some of the cold gas to heat up and become hot gas. It also heats hot gas up and cause some of it to flow out of galaxy in a wind. This phenomenon is called supernova feedback (SN feedback). This process is the main agent controlling the overall efficiency of star formation (Larson, 1974; White and Rees, 1978; Dekel and Silk, 1986; Benson et al., 2003). Therefore, we assume that SN feedback injects gas from cold gas to hot gas and, in addition, can transfer hot gas to the ejected gas reservoir which it may or may not be reincorporated at later time, depending on the mass of the halo hosting the galaxies (Guo et al., 2011).

We follow Henriques et al. (2015) to estimate the energy from supernovae

$$\Delta E_{\rm SN} = \epsilon_{\rm halo} \times \frac{1}{2} \Delta m_{*} v_{\rm SN}^{2}, \tag{A.14}$$

where $\frac{1}{2} v_{\rm SN}^{2}$ is the mean energy per unit mass of formed stars and the efficiency

$$\epsilon_{\rm halo} = \eta \times \left[0.5 + \left(\frac{v_{\rm max}}{v_{\rm ejected}} \right)^{-\beta_2} \right], \tag{A.15}$$

as η and β_2 are parameters. The mass of cold gas reheated by star formation and added into hot atmosphere is assumed to be

$$\Delta m_{\rm reheat} = \epsilon_{\rm disk} \Delta m_{*} \tag{A.16}$$

where the efficiency is

$$\epsilon_{\rm disk} = \epsilon \times \left[0.5 + \left(\frac{v_{\rm max}}{v_{\rm reheat}}\right)^{-\beta_1}\right], \tag{A.17}$$

as ϵ and β_1 are parameters. We assume the energy required for reheating:

$$\Delta E_{\rm reheat} = \frac{1}{2}\Delta m_{\rm reheat} v_{\rm 200c}^2. \tag{A.18}$$

We assume that the required energy for reheating $\Delta E_{\rm reheat} = \frac{1}{2} m_{\rm reheat} v_{\rm 200c}^2$. If $\Delta E_{\rm reheat} > \Delta E_{\rm SN}$, the reheated mass will be set at the maximum value of $\Delta m_{\rm reheat} = \Delta E_{\rm SN} / \left(\frac{1}{2} v_{\rm 200c}^2\right)$. If $\Delta E_{\rm reheat} \leq \Delta E_{\rm SN}$, the leftover energy from reheating is used to eject a mass of $\Delta m_{\rm ejected}$ from hot gas into the ejected gas reservoir, where

$$\frac{1}{2}\Delta m_{\rm ejected} v_{\rm 200c}^2 = \Delta E_{\rm SN} - \Delta E_{\rm reheat}. \tag{A.19}$$

The discussions of this topic can be found in Henriques et al. (2015).

A.6 Reincorporation of ejected gas

We follow Henriques et al. (2013) to assume that the increase rate of hot gas taken from ejected phase:

$$\dot{m}_{\rm ejected} = -\frac{m_{\rm ejected}}{t_{\rm reinc}}, \tag{A.20}$$

where the reincorporation time-scale:

$$t_{\rm reinc} = \gamma \frac{10^{10}\ {\rm M}_\odot}{M_{\rm 200c}}, \tag{A.21}$$

where γ is a parameter.

A.7 Metal enrichment

We follow Henriques et al. (2015) to assume that, when stars die, they release newly formed heavy elements into surrounding aside from mass and energy. Star formation process converts a fixed proportion of material of raw material into metals. The newly formed metals are mixed instantly into the cold gas.

A.8 Satellite galaxies

We classify galaxies into three types according to their relationship with dark matter haloes (Guo et al., 2011):

- **Type 0 galaxies (main galaxies)** are the central galaxies of the main subhaloes in FoF groups and can be considered as the main galaxies of their FoF groups.
- **Type 1 galaxies (satellites)** are at the centres of non-dominant subhaloes.
- **Type 2 galaxies (orphan galaxies)** are the galaxies which are not in any dark matter subhalos any more.

All galaxies are born as type 0. When the subhaloes they reside in fall into groups or clusters, they become type 1. Later, when the dark matter subhaloes they reside in merge into the main subhaloes or they cannot be resolved, they will become type 2 galaxies.

When galaxies become satellites of larger systems, they will be influenced by tidal force. This effect will remove materials and modify the structure and evolution of the galaxies. We follow the processes from Guo et al. (2011) and Henriques et al. (2015) to model this effect.

A.8.1 Tidal and ram-pressure stripping

When a subhalo falls in to a larger system, the tidal force will begin to remove its dark matter (De Lucia et al., 2004a). We follow Guo et al. (2011) to assume that there is no baryonic infall in this subhalo and its hot gas will be stripped away as follows:

$$\frac{m_{\rm hot}(R_{\rm tidal})}{m_{\rm hot,infall}} = \frac{M_{\rm DM}}{M_{\rm DM,infall}} \tag{A.22}$$

where the limiting radius, $R_{\rm tidal}$, is given by a simple isothermal model,

$$R_{\rm tidal} = R_{\rm DM,infall} \times \left(\frac{M_{\rm DM}}{M_{\rm DM,infall}}\right). \tag{A.23}$$

$M_{\rm DM,infall}$ and $R_{\rm DM,infall}$ are $M_{\rm 200c}$ and $R_{\rm 200c}$ of the halo just before the infall. $m_{\rm hot,infall}$ is the hot gas mass of the main galaxy prior to the infall. $M_{\rm DM}$ and $m_{\rm hot}$ are $M_{\rm 200c}$ of the halo and the hot gas of the main galaxies at the current time respectively. Thus, the tidal stripping will remove all hot gas once the subhalo is not resolved and the galaxy becomes an orphan.

Ram pressure can strip hot gas from a satellite as soon as it enters the virial radius of massive host (here we use a threshold of 10^{14} $\rm M_\odot$). At a distance $R_{\rm r.p.}$ from the centre of the satellite, its self-gravity is approximately equal to the ram pressure:

$$\rho_{\rm sat}(R_{\rm r.p.})v_{\rm sat}^2 = \rho_{\rm host}(R)v_{\rm orbit}^2, \tag{A.24}$$

where $\rho_{\rm sat}(R_{\rm r.p.})$ is the hot gas density of the satellite at radius $R_{\rm r.p.}$, $v_{\rm sat}$ is the virial velocity of the subhalo which we assume to be constant as it orbits around the main halo,

$\rho_{\rm host}(R)$ is the hot gas density of the host at the distance R from the centre of the potential well to the satellite, $v_{\rm orbit}$ is the orbital velocity of the satellite, which we assume circular orbiting. The densities here are assume to be $\rho \propto r^{-2}$ according to a simple isothermal model.

A.8.2 Tidal disruption

Following Guo et al. (2011) to assume that disruption only occur on Type 2 galaxies which has depleted their hot gas. The condition for disruption is, if the baryon (cold gas + stellar mass) density within the half mass radius, $\rho_{\rm sat} = m_{\rm sat}/R^3_{\rm sat,half}$, is less than the dark matter density of the main halo within the pericentre of the satellite's orbit, $\rho_{\rm DM,host} = M_{\rm DM,host}(R_{\rm peri})/R^3_{\rm orbit}$, the satellite is completely disrupted. Its stars will be added to the intracluster light (ICL) and its cold mass will be added to the hot gas of the central galaxy. We assume the exponential surface density profiles for stars and ICL and a surface density proportional to $r^{1/4}$ for hot gas.

By assuming conservation of energy and angular momentum and a singular isothermal potential for the orbit, $\phi(R) = v^2_{\rm 200c} \ln R$, where R is the distance from the centre of halo to the satellite. The orbital pericentre can be calculated as

$$\left(\frac{R}{R_{\rm peri}}\right)^2 = \frac{2\ln R/R_{\rm peri} + (v/v_{\rm 200c})^2}{(v_t, v_{\rm 200c})^2}, \tag{A.25}$$

where v and v_t are the total and tangential velocities of the satellite respect to its host halo centre.

A.9 Mergers and bulge formation

A.9.1 Positions and velocities of orphans

After the disruption, the positions and velocities of Type 2 galaxies are linked to their most bound dark matter particles just before they were disrupted. As soon as a disruption happens, the satellite will be assumed to spiral into the central object through the dynamical friction. We estimate the time taken in this process by using Binney and Tremaine (1987) formula:

$$t_{\rm friction} = \alpha_{\rm friction} \frac{v_{\rm 200c} r^2_{\rm sat}}{G m_{\rm sat} \ln \Lambda}, \tag{A.26}$$

where $m_{\rm sat}$ is the total mass of the satellite, $\ln \Lambda = \ln(1 + M_{\rm 200c}/m_{\rm sat})$ is the Coulomb algorithm and $\alpha_{\rm friction} = 2.4$ (De Lucia and Blaizot, 2007).

As usual, we follow Guo et al. (2011) to model the decay of satellite's orbit due to dynamical friction by placing the Type 2 galaxy not at at a position whose (vector) offset from the central galaxy is reduced from that of the particle by a factor of $(1 - \Delta t/t_{\rm friction})$ where Δt is the time since the disruption. The velocity of the orphan galaxy is set equal to that of the most bound particle. $t_{\rm friction}$ is the total time the orphan takes to merge into the central galaxy.

A.9.2 Bulge formation

We separate mergers into two categories. A major merger is when the ratio between the total baryonic mass of the smaller galaxy and the total baryonic mass of the larger galaxy exceeds a fractional value, $0 \leq R_{\rm merger} \leq 1$. In this work, we set $R_{\rm merger} = 0.1$ (Henriques et al., 2015). In a major merger, the discs of the two progenitors are destroyed and all their stars become part of the bulge of the descendent together with the star formed during the merger. For a minor merger, the disc of the larger progenitor survives and accretes the cold gas from the the smaller galaxy. The forming bulge receives all star from the smaller galaxy while the stars formed during the merger stay in the disc of the descendent (see A.9.3). In both cases, cold gas is transferred to the central black hole (see Section A.10).

By assuming energy conservation and virial theorem, the change in sizes in both minor and major mergers:

$$\frac{m_{\rm new,bulge}^2}{R_{\rm new,bulge}} = \frac{m_1^2}{R_1} + \frac{m_2^2}{R_2} + 2\alpha_{\rm inter}\frac{m_1 m_2}{R_1 + R_2}, \tag{A.27}$$

where m_1 and m_2 are the baryonic masses of the two progenitors. $\alpha_{\rm inter}$ is coefficient for binding energy in the orbit relative to that in the individual system. It is set to 0.5 and have a good agreement with SDSS data (Guo et al., 2011). When either of the progenitors is a composite disc-and-bulge system, its half-mass radius is calculated assuming an exponential disc and an $r^{1/4}$-law bulge.

The disk instabilities is an important factor in bulge growth. They transport materials from the disc to the bulge. This happens in galaxies where self-gravity of the disc domimates the gravitational effects of the bulge.

The criterion for disc instability is

$$v_{\rm max} < \sqrt{\frac{G m_{*,\rm d}}{3 R_{*,\rm d}}}, \tag{A.28}$$

where $m_{*,\rm d}$ and $R_{*,\rm d}$ are the stellar mass and exponential scale-length of the stellar disc and $v_{\rm max}$ is the maximum circular velocity of the host dark matter halo hosting the galactic disc Guo et al. (2011).

When the eq. A.28 is satisfied, some of stellar mass from the disc is transferred to the bulge to stabilise the disc. We estimate the transferred mass,

$$\Delta m_* = 2\pi\Sigma_{*,0}\left[R_{*,\mathrm{d}} - (R_\mathrm{b} + R_{*,\mathrm{d}})\exp\left(\frac{-R_\mathrm{b}}{R_{*,\mathrm{d}}}\right)\right], \tag{A.29}$$

where $\Sigma_{*,0}$ and $R_{*,\mathrm{d}}$ is the central surface density and the scale-length of the unstable disc respectively. We assume that the angular momentum of the transferred material is negligible so $R_{*,\mathrm{d}}$ will increase due to the conservation of angular momentum (Guo et al., 2011).

A.9.3 Star burst from mergers

We follow Somerville et al. (2001) to estimate the star formation during a merger,

$$m_{*,\mathrm{burst}} = \alpha_{\mathrm{SF,burst}} m_\mathrm{cold}\left(\frac{m_1}{m_2}\right)^{\beta_{\mathrm{SF,burst}}} \tag{A.30}$$

where $m_1 < m_2$ are the baryonic masses of the two galaxies and m_cold is their total cold gas mass. $\alpha_{\mathrm{SF,burst}}$ and $\alpha_{\mathrm{SF,burst}}$ are free parameters.

A.10 Black holes

A.10.1 Black hole growth - quasars

When galaxies merge, their cold gas components are pulled into the bulge. A large amount of gas will be in the near the inner region of the galaxy where it might form a black hole or be accreted into the existing black hole. If both progenitor galaxies contain pre-existing black holes, they will be combined in the descendent.

The amount of gas accreted into the quasar is taken to be:

$$\Delta m_{\mathrm{BH,Q}} = \frac{f_\mathrm{BH} m_\mathrm{cold}(m_\mathrm{sat}/m_\mathrm{cen})}{1 + (v_\mathrm{BH}/v_{\mathrm{200c}})^2}, \tag{A.31}$$

where m_cen and m_sat are total baryonic masses of the central galaxy and the merging satellite respectively, f_BH and v_BH are adjustable parameters, m_cold is the total cold gas mass of the system, and v_{200c} is the virial velocity of the central halo.. The final mass of the black hole at the centre of the descendent halo $m_\mathrm{BH} = m_{\mathrm{BH},1} + m_{\mathrm{BH},2} + m_{\mathrm{BH,Q}}$ where $m_{\mathrm{BH},1}$ and $m_{\mathrm{BH},2}$ are the mass of the two progenitor black holes.

A.10.2 Black hole radiative feedback

We assume that the central supermassive black holes accrete materials from hot gas atmosphere continuously. This produce the feedback by injecting energy into the hot gas.

We follow Henriques et al. (2015) to approximate the accretion rate:

$$\dot{m}_{\rm BH} = k_{\rm AGN} \left(\frac{m_{\rm hot}}{10^{10}\ {\rm M}_\odot} \right) \left(\frac{m_{\rm BH}}{10^{8}\ {\rm M}_\odot} \right). \tag{A.32}$$

The accretion will produce relativistic jets which transfer energy into the hot gas. The energy input rate is

$$\dot{E}_{\rm radio} = \eta \dot{m}_{\rm BH} c^2, \tag{A.33}$$

where $\eta = 0.1$ is an efficiency parameter and c is the speed of light. This energy suppresses the cooling from hot gas to cold disc. The effective cooling rate becomes

$$\dot{m}_{\rm cool,eff} = max \left[\dot{m}_{\rm cool} - 2\dot{E}_{\rm radio}/v^2_{\rm 200c}, 0 \right] \tag{A.34}$$

Here, we assume that if of the cooling flow also cuts off the supply of gas to the black hole, so that heating of the hot atmosphere beyond this point is not possible.

A.11 Stellar populations synthesis

Stellar population synthesis models are a very important part of galaxy formation theory as they link the masses, ages and metallicities predicted for stars to the observable emission at various wavelengths. We use the publicly released but still unpublished Charlot & Bruzual (2007) because there is a number of authors suggest that it has a better agreement with the observations than Bruzual and Charlot (2003) (Henriques et al., 2011, 2012, 2015; Tonini et al., 2009, 2010; Fontanot and Monaco, 2010; Gonzalez-Perez et al., 2014).

A.12 Dust model

Since dust is able to significantly absorb optical/UV photons, it will have a significant effect on the observation of star-forming galaxies which are expected to be rich in dust. Therefore, dust-dominated galaxies will generally have red colours even if they are actively star-forming. For that reason, we summarise the dust model of Henriques et al. (2015) here: more detailed description can be found in Section 1.14 of the supplementary material in that paper.

By considering dust extinction separately for the diffuse interstellar medium (ISM) and for the molecular birth clouds within which stars form. The optical depth of dust as a function of wavelength is computed separately for each component. Then, we combine them as the descriptions below to compute the total extinction of the relevant populations.

A.12.1 Extinction by the ISM

The optical depth of diffused dust in galactic discs is assumed to depend on wavelength as

$$\tau_\lambda^{\rm ISM} = (1+z)^{-1} \left(\frac{A_\lambda}{A_{\rm V}}\right)_{Z_\odot} \left(\frac{Z_{\rm gas}}{Z_\odot}\right)^s \times \left(\frac{\langle N_H \rangle}{2.1 \times 10^{21}\ {\rm atom\ cm^{-2}}}\right), \tag{A.35}$$

where

$$\langle N_H \rangle = \frac{m_{\rm cold}}{1.4\pi m_p (a R_{\rm gas,d})^2} \tag{A.36}$$

is the mean column density of hydrogen, 1.4 accounts for the presence of helium, $R_{\rm gas,d}$ is the cold disc scale-length and $a = 1.68$. Following the results in Guiderdoni and Rocca-Volmerange (1987), the extinction curve in Equation A.35 depends on the gas metallicity and is based on the observation in local Universe: $s = 1.35$ for $\lambda < 2000$Åand $s = 1.6$ for $\lambda > 2000$Å. The extinction curve for solar metallicity, $(A_\lambda/A_{\rm V})_{Z_\odot}$, is taken from Mathis (1983).

A.12.2 Extinction by molecular birth clouds

This second source of extinction affects only young stars, for which we assume a lifetime of 10 Myr. The relevant optical depth is taken to be

$$\tau_\lambda^{BC} = \tau_\lambda^{\rm ISM} \left(\frac{1}{\mu} - 1\right) \left(\frac{\lambda}{5500\text{Å}}\right)^{-0.7}, \tag{A.37}$$

where μ is given by a random Gaussian deviate with mean 0.3 and standard deviation 0.2, truncated at 0.1 and 1.

A.12.3 Overall extinction curve

In order to get the final overall extinction, every galaxy is assigned an inclination, θ, given by the angle between the disc angular momentum and the z-direction of the simulation box, and a "slab" geometry is assumed for the dust in the diffuse ISM. For sources that are uniformly distributed within the disc then the mean absorption coefficient is

$$A_\lambda^{\rm ISM} = -2.5 \log_{10} \left(\frac{1 - \exp^{-\tau_\lambda^{\rm ISM} \sec\theta}}{\tau_\lambda^{\rm ISM} \sec\theta}\right), \tag{A.38}$$

Emission from young stars embedded within birth clouds is subject to an additional extinction of

$$A_\lambda^{\rm BC} = -2.5 \log_{10} \left(\exp^{-\tau_\lambda^{\rm BC}}\right). \tag{A.39}$$

Appendix B

The tree data format

In order to facilitate comparison and use of merger tree data, it is our intention to define in a future paper a common merger tree data format. This should make provision for: required minimal data to define a merger tree; desired fields to ease use; and the ability to include optional additional data that may prove useful. At the time of writing (prior to the Sussing Merger Trees Workshop) that format had not been defined and so we restrict ourselves to outlining here the minimal data format that was used for the work described in this paper.

We supplied each participant in the tree comparison project with a list of haloes, together with their properties (as described in Section 6.3) and an inclusive list of particle IDs. Each halo had a identifier (halo ID) that was unique across snapshots.

We required participants to return their results in the ascii format described in Table B.1, where there is an entry for each halo. That contains enough information for us to be able to reconstruct the merger trees and, in conjuntion with the original halo list, to follow the growth of haloesover time.

Table B.1: The ASCII data format that participants were asked to use to return their merger tree results.

Information to be returned	Notes
FormatVersion	= 1 – an integer indicating the format version
Description	Name of code, version/date of generation; max 1024 characters
Nhalo	Total number of haloes specified in this file
HaloID_1, N_1	Halo's ID and number of direct progenitors
Progenitor_1	Halo ID of main progenitor of halo HaloID_1 (where $N_1 > 0$)
Progenitor_2	Halo IDs of other progenitors of halo HaloID_1
...	...
Progenitor_{N_1}	Halo ID of last progenitor of halo HaloID_1
...	...
$\text{HaloID}_{\text{Nhalos}}$, N_{NHalo}	Halo's ID and number of direct progenitors
$\text{Progenitor}_{\text{NHalo}}$	Halo ID of main progenitor of halo $\text{HaloID}_{\text{NHalo}}$ (where $N_{\text{NHalo}} > 0$)
Progenitor_2	Halo IDs of other progenitors of halo $\text{HaloID}_{\text{NHalo}}$
...	...
$\text{Progenitor}_{N_{\text{NHalo}}}$	Halo ID of last progenitor of halo $\text{HaloID}_{\text{NHalo}}$
END	String 'END' indicating the last line of the output file